FACULTEIT ECONOMIE EN
BEDRIJFSWETENSCHAPPEN

KU LEUVEN

Economics of
Power Transmission Reliability

Proefschrift voorgedragen tot
het behalen van de graad van
Doctor in de
Economische Wetenschappen

door

Marten Ovaere

Number 564 2017

Doctoral Committee

Advisor:
Prof. dr. Stef Proost (KU Leuven)

Members:
Prof. dr. Fridrik Már Baldursson (Reykjavik University, University of Oslo)
Prof. dr. Thomas-Olivier Léautier (Toulouse School of Economics, EDF)
Prof. dr. Guido Pepermans (KU Leuven)
Prof. dr. Dirk Van Hertem (KU Leuven)
Prof. dr. Frank Verboven (KU Leuven)

Chair:
Prof. em. Willem Moesen (KU Leuven)

Acknowledgements

Two roads diverged into the woods and I took one filled with wonderful people, and that has made all the difference.

First of all, Stef Proost, who convinced me to go down this road and who gave me the autonomy and trust I needed. I have learned a lot from him, both explicitly and implicitly. His patience, motivation, and economic insight are an example. He taught me that there is no royal road to a good paper and gave me important directions through small nudges.

My sincere thanks also go to the members of the doctoral committee for their time and effort invested in reading my work and for providing me with constructive comments. Fridrik Már Baldursson, for his serene and kind way of managing me and our project; Dirk Van Hertem, for his extensive knowledge of transmission reliability; Guido Pepermans and Frank Verboven, for their detailed comments and new insights; and Thomas-Olivier Léautier, for enabling my path to pass through Toulouse.

During this five-year track, I was fortunate to have met numerous brilliant people within the Department of Economics. First of all, I would like thank my colleagues for all the interesting discussions over lunch, tea, soup or beer. It was a pleasure sharing my pit stops with you. A special thanks to my 'compagnons de route' Pieter Jan, Geert and Joris, who were always available for research and non-research related conversations. Also Karla, Heidi, Pascale and Andras deserve a special mention for their kind assistance and for making sure my administration stayed on track.

Along the road, I have also met many great people outside of the Department, most of whom share my passion for energy. I would like to thank my 'energy-colleagues' from engineering, especially Arne and Kenneth, with whom I have driven through energy markets and Utopian lands. Someone who also deserves a special mention is Evelyn, an outstanding and quick-witted co-author, with excellent modelling skills. I was also fortunate to get to know Felix, Nicolas and Xavier, who warmly welcomed me at Toulouse School of Economics. Lastly, the monthly detour in sustainability and the multidisciplinary discussions with the people from YouRSS were always very enriching. It was great meeting all these like-minded and passionate people!

I consider myself very lucky to still be surrounded by so many lifelong friends. My thanks to the '6B' friends for the lovely dinners, weekends and walks. To the 'Harmonie' friends for clearing my head with music on Friday evenings. And to the 'Leuven Partysquad' friends for the dinners, holidays, discussions, adventures, and for the many 'one for the road's'.

I am incredibly fortunate to have such a wonderful family around me. I would like to thank them for all their love and encouragement. First of all, my parents who supported me in all my pursuits and who taught me the important values of understanding, respect and kindness. You are truly an example. Next, my siblings and their partners, for being great people and for enduring my arguments at the kitchen table. Lastly, my in-laws for welcoming me into their family.

Finally, I am extremely lucky to have such a magnificent co-driver next to me and such a fabulous little daughter in the back seat (who made the last miles more challenging but also much more fulfilling). Liesbet, I am immensely grateful for your unconditional love, your support, your understanding and for just being your lovely self. We are a great team. I have enjoyed every second of our ride together and I look forward to our joint road ahead.

I hope to see you all again somewhere down the road!

Marten
Leuven, June 2017.

Contents

Introduction

Electricity is the backbone of modern society. We want electricity to be available at all times, because without it, lights go out, appliances stop working, and factories shut down. Unfortunately, the reliability of our electricity supply is continuously challenged by adverse weather, loop flows, forecast errors, variable generation and consumption, and unplanned outages of lines, transformers, generation plants and large loads. As the cost of a major interruption or blackout is very large, network operators do their utmost to achieve a high reliability level, but a completely reliable electricity supply comes at an infinite cost.

Traditionally, the energy sector has emphasized reliability rather than the costs of achieving this reliability. However, considering the large cost of both electricity provision and electricity interruptions, it is of paramount importance to determine the correct reliability level.

This doctoral thesis addresses this question by studying the fundamental trade-off between the value of reliability and its cost. The six chapters of this thesis focus on different aspects of power system reliability:

1. Optimal transmission reliability.
2. Choosing a reliability criterion.
3. Optimal regulation of reliability.
4. Cooperation in generation reliability.
5. The importance of the value of lost load in reliability management.
6. The effect of reliability management on inequality of reliability.

Context and motivation

A power system consists of four primary components:

(A) **Generation**: the process of generating electricity from coal, gas, oil, nuclear fuel, wind, sun, hydro, etc.

(B) **Transmission**: the high-voltage network of cables, lines and transformers that transports electricity from generation plants to the distribution network.

(C) **Distribution**: the low-voltage network that carries electricity from the transmission system to individual consumers.

(D) **Retail**: the final sale of electricity to end-consumers.

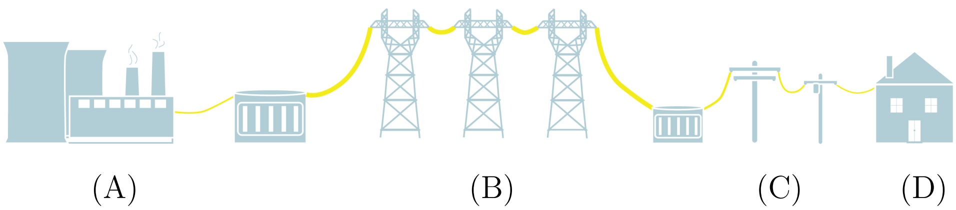


Figure 1: The four primary components of power systems. (Source: Institute for Energy Research)

For decades all four primary components were integrated within a single public electric utility that was responsible for generating electricity, transporting it from generation plants to load centers, distributing it and selling it to individual consumers. From the 1980s, in the wake of a deregulation movement that had already affected airlines, transportation and the supply of gas, the vertically integrated and centrally planned electricity system was gradually replaced by a system that relied on markets and different market players (Kirschen and Strbac, 2004a). Although the exact design of these deregulated or liberalized markets differs between countries, they have the following in common:

- Separation of the four primary components.
- Introduction of competition in generation and retail. The existing portfolio of generation plants is split in multiple competing companies and free entry is allowed for companies that meet certain requirements.

- Creation of separate transmission and distribution companies to manage the monopoly networks.
- Allowing consumers to choose between competing retailers.
- Establishment of a national regulatory authority that monitors both the competing companies and the monopoly networks.
- Creation of markets where generators, retailers, large consumers and the transmission system operator make transactions.
- Some of the new companies are still publicly owned, while others are private.

This thesis mainly focuses on the transmission system operator (TSO), the entity responsible for the transmission network. In addition to managing its own transmission network, the TSO is also responsible for maintaining a continuous balance between generation and load. This is a difficult task for three reasons. First, electricity demand varies on a daily, weekly or even seasonal basis, it is difficult to predict, and it does not (yet) react to the real-time price of electricity. Second, the increased penetration of renewables also makes generation variable and uncertain. Third, network components, generation plants and large loads can unexpectedly fail.

To ensure that generation and load are always balanced, despite the above problems, a TSO takes many reliability actions. Even if this balance is evidently a real-time issue, these actions are taken at different time horizons before real-time operation:

- **System expansion**: construction, upgrading, replacement, retrofitting or decommissioning of assets like AC or DC high-voltage transmission lines, substations, shunt reactors, phase-shifting transformers, etc.
- **Asset management**: monitoring the health status of network components, planning maintenance activities, repairing the components in case of failure, etc.
- **Operational planning**: congestion management, system protection, reserve provision, preventive actions, voltage control, decisions on outage executions, etc.
- **Real-time operation**: corrective actions, activation of reserves, reliability assessment, etc.

By taking these actions, a TSO prepares its electricity system for a set of possible operating states and possible contingencies. For example, a combination of stormy weather, very high wind production and the failure of an important transmission line. However, a TSO can not prepare its system for every possible operating state or every possible contingency. For example, securing the system against the simultaneous failure of three nuclear power plants and four transmission lines would require a too high cost of respectively generation reserves and transmission capacity. This brings up the question what constitutes an appropriate or even optimal level of reliability. Answering this require us to analyze the trade-off between costs and benefits of reliability actions.

This trade-off, however, is not at the heart of a TSO's reliability management. Currently, TSOs do not explicitly target a certain reliability level. Instead, their reliability actions aim to meet the N-1 reliability criterion, which states that an unexpected outage of a single system component (lines, transformers, generation plants, large loads, etc.) can not result in a loss of load. That is, when a single system component fails, the transmission system should still be able to accommodate all flows without load shedding. This also implies that, following the N-1 reliability criterion, a simultaneous failure of multiple system components could require load shedding to avoid a black out.

The straightforward N-1 reliability criterion has led to satisfactory results in the last decades, but is increasingly challenged by researchers and decision makers. They argue that the N-1 criterion is not able to efficiently meet the current challenges of the electricity system, such as uncertain and variable demand and supply, decentralized decision makers, highly interconnected networks, difficulties in building new lines and a general trend towards a more efficient use of the transmission system. The N-1 reliability criterion is intuitive and easy to understand, but lacks transparency and flexibility, and ignores the economic trade-off between costs and benefits. This trade-off is needed to determine optimal reliability actions and the optimal reliability level.

The socio-economic cost of having a too high or too low level of reliability are substantial. As the cost of reliability actions is paid by end-consumers of electricity, increasing the reliability level raises their cost, which is already a considerable part of household expenditures. For example, in the United Kingdom, households spend between 1.4% (richest 10%) and 4.8% (poorest 10%) of their income on electricity (OFGEM, 2017a). On the other hand, the cost of interruptions is equally large. A one-day blackout could amount to about 0.5% of a country's GDP (Kirschen, 2002) and could, in addition, lead to societal consequences such as injuries, diseases and deaths.

This dissertation analyzes the question of optimal transmission reliability. To tackle this problem, we need to address several questions: What is the set of possible operating states and contingencies to consider? What is the probability of different operating states and contingencies? What is the cost of interruptions? What are possible reliability actions? What is the information available to the TSO? What is a TSO allowed to do? How is a TSO regulated?

Outline and contributions

All six chapters of this thesis deal with the fundamental trade-off between reliability costs and interruption costs. They combine analytical models with case studies, small numerical illustrations or larger simulation models. The target audience are economists and engineers studying power systems, as well as policy makers and transmission system operators. In a nutshell, the six chapters of this thesis cover the following topics:

Chapter I: Optimal Electricity Transmission Reliability: Going Beyond the N-1 Criterion. In the presence of transmission outages, uncertain demand and variable renewable supply, network operators keep a reliability margin to avoid interruptions and black-outs. The reliability margin is presently determined by the N-1 reliability criterion. Our analytical model defines the optimal reliability margin by balancing congestion costs and interruption costs. This leads to more efficient use of transmission capacity and to smaller investment needs than with the N-1 criterion. A numerical illustration shows the net benefits of the new reliability criterion.

This chapter is joint work with Stef Proost. It has been presented at YEEES (Leuven, Belgium, 2014), CIES (Cologne University, Germany, 2015), the BAEE meeting (UCL, Belgium, 2015), and the IAEE international conference (Bergen, 2016), where it has won a Student Best Paper Award. An earlier version can be found in:

- Ovaere, M., and Proost, S. (2016). Electricity transmission reliability: The impact of reliability criteria. KU Leuven Department of Economics Discussion Paper Series, 16.21.

Chapter II: A Multi-Dimensional Analysis of Reliability Criteria: From Deterministic N-1 to a Probabilistic Approach. This chapter proposes a classification of reliability criteria for power systems based on four characteristics: (i) the set of considered system states, (ii) the objective function, (iii) the allowed real-time actions and (iv) optional non-technical constraints. Because selecting a reliability criterion involves a trade-off between multiple objectives, this chapter suggests the use of five performance indicators to evaluate reliability criteria: (i) expected total cost, (ii) reliability level, (iii) inequality between consumers, (iv) data needs and availability and (v) ease of use. A case study for a five-node test system illustrates this multi-dimensional analysis of six reliability criteria that range from the deterministic N-1 criterion to a full probabilistic criterion that aims to minimize expected total cost. This analysis finds that moving from deterministic to probabilistic reliability criteria decreases expected total cost, but unreliability and inequality increase. The largest savings of expected total cost are due to a trade-off between preventive and corrective actions. A smaller portion of savings is due to the trade-off between preventive and curtailment actions. Limits on individual or aggregate unreliability levels decrease unreliability and inequality, but increase expected total costs when compared to a fully probabilistic approach.

This chapter is joint work with with Evelyn Heylen, Stef Proost, Geert Deconinck and Dirk Van Hertem.

Chapter III: Cost and Quality Regulation of a Monopolist. This chapter studies the effect of linear cost and quality regulation of a monopolist on its cost-reducing effort and its provided quality level. The model differs from the earlier literature on quality regulation by focusing on the monopolist's cost function instead of on the demand curve. As a result, we analyze the effect of regulation on quality and cost-reducing effort instead of quality and quantity. The analysis shows that both quality and effort increase with the power of the quality incentive, while the effect of the power of the cost incentive is ambiguous. Next, introducing uncertainty, the power of the cost incentive and quality incentive should optimally be equal and below one. Last, we compare our hybrid regulation to pure rate-of-return and price-cap regulation and analyze case studies in electricity, gas and water.

This chapter has been presented at YEEES (The University of Edinburgh, UK, 2016) and the FAEE Student Workshop (MINES ParisTech, France, 2017).

Chapter IV: Cross-Border Exchange and Sharing of Generation Reserve Capacity. This chapter develops a stylized model of cross-border balancing. We distinguish three degrees of cooperation: autarky, reserves exchange and reserves sharing. The model shows that TSO cooperation reduces costs. The gains of reserves exchange increase with cost asymmetry and the additional gains of reserves sharing decrease with correlation of real-time imbalances. Based on actual market data of reserves procurement of automatic frequency restoration reserves in Belgium, France, Germany, the Netherlands, Portugal and Spain, we estimate the efficiency gains of exchange to be around €60 million per year and of sharing to be around €150 million per year. The model also shows that voluntary cross-border cooperation could be hard to achieve, as TSOs do not necessarily have correct incentives.

This chapter is joint work with M. Baldursson, Ewa Lazarczyk and Stef Proost. It has been presented at the IAEE international Conference (Bergen, 2016). Earlier versions can be found in:

- Baldursson, F. M., Lazarczyk, E., Ovaere, M., and Proost, S. (2016). Cross-border exchange and sharing of generation reserve capacity, CREE Working paper 14/2016.
- Baldursson, F. M., Lazarczyk, E., Ovaere, M., and Proost, S. (2016). Multi-TSO system reliability: Cross-border balancing. IEEE International Energy Conference, ENERGYCON 2016.

Chapter V: How Detailed Value of Lost Load Data Impact Power System Reliability Decisions: a Trade-off Between Efficiency and Equity. The value of lost load (VOLL) is an essential parameter for transmission system reliability management. It represents the cost of unserved energy of electricity interruptions. Various studies have estimated this parameter for different countries and more recently, for different interruption characteristics – such as interruption duration, time of interruption and interrupted consumer. However, most applications only use one uniform VOLL. Our theoretical analysis shows that using more-detailed VOLL data leads to better-informed transmission reliability decisions. To illustrate this, we estimate the efficiency gains of including consumer and time characteristics in short-term transmission reliability management using VOLL data from Norway, Great Britain and the United States. Depending on the VOLL data and the method of demand curtailment, our five-node network indicates efficiency gains up to 43%. However, increased efficiency leads to decreased equity. Striking the balance between these opposing objectives is crucial for social acceptance.

This chapter is joint work with with Evelyn Heylen, Stef Proost, Geert Deconinck and Dirk Van Hertem. It has been presented at EnergyVille (Belgium, 2016) and the CIGRE C5 Mirror Meeting (Belgium, 2016). An earlier version can be found in:

- Ovaere, M., Heylen, E., Proost, S., Deconinck, G., and Van Hertem, D. (2016). How detailed value of lost load data impact power system reliability decisions: a trade-off between efficiency and equality. KU Leuven Department of Economics Discussion Paper Series, 16.26.

Chapter VI: An Inequality Indicator of Power System Reliability. Power system decisions not only affect the overall reliability level, but also the reliability level of individual load points and consumers depending on their location and consumer characteristics. A Gini-based indicator is proposed that summarizes the inequality of power system reliability between different entities, such as consumers, nodes or regions. The indicator expresses the perceived fairness of the reliability level. Because fairness contributes to social acceptance of power system decisions, the inequality indicator can help decision makers to assess the effect of power system decisions on inequality of reliability and to take appropriate actions to decrease this inequality. The use of the inequality indicator is illustrated using two case studies: (i) the Belgian load-shedding plan for generation adequacy and (ii) the shift from deterministic to probabilistic short-term power system reliability management.

This chapter is joint work with with Evelyn Heylen, Stef Proost, Geert Deconinck and Dirk Van Hertem.

Chapter I

Optimal Electricity Transmission Reliability: Going Beyond the N-1 Criterion

I.1 Introduction

Reliability of electricity supply is of paramount importance to our society. Without electricity, lights go out, appliances stop working, and factories shut down. A transmission system operator (TSO) is entrusted with the task of safeguarding our supply. Reliability management affects all its decisions, from long-term system development to short-term operational planning and system operation.

Despite its importance, the economic literature has not given much attention to electricity transmission reliability (Joskow, 2006).[1] Many papers study economic transmission investment (e.g. Turvey (2006); van der Weijde and Hobbs (2012); Doucet et al. (2013); Pozo et al. (2013)) and its regulation (e.g. Léautier (2000); Rosellón (2007); Rosellón and Weigt (2011)), but

[1] Joskow (2006, p.12) states that ”neither reliability transmission investments nor the interrelationship between reliability criteria and economic parameters are given much attention in the literature on competitive electricity markets.”

leave reliability issues aside. Considering reliability, however, is important to better understand transmission capacity investment and the use of this transmission capacity. For example, many European countries aim for more transmission investment to better cope with renewable energy integration and to lower wholesale electricity prices.[2] But, because of the costs of transmission investments and the difficulties in building new lines (Cohen et al., 2016), the question is whether a more efficient use of current transmission capacity is a better alternative. Addressing this question requires us to think about electricity transmission reliability.

This paper considers the possibility of transmission line outages, generation outages, uncertainty of demand, and variable renewable supply. The paper shows that TSOs use less transmission capacity than is installed, because of reliability concerns. As a result, the effect of transmission investment on reliability depends on the TSO's system operation, which is managed by its reliability criterion. Currently all TSOs use the N-1 reliability criterion or some variant. This deterministic criterion states that an unexpected outage of a single system component may not result in a loss of load. We show that the N-1 reliability criterion is suboptimal since it only depends on the topology and the use of the network, not on network conditions and economic parameters. In addition, the model provides insight into the trade-off between reliability and congestion. It shows that the N-1 criterion requires the TSO to achieve an exogenous reliability level and minimize congestion costs, while optimally the reliability level and the congestion cost are determined endogenously (Hogan et al., 2010).

To our knowledge, an analytical economic model of electricity transmission reliability does not yet exist. Some economic papers have, however, discussed electricity transmission reliability and reliability criteria. Blumsack et al. (2007) study the relationship between congestion and reliability, but limit their focus to a Wheatstone network. They find that in this particular network topology, investing in a transmission line to increase reliability could increase congestion. By contrast, we will show in this paper that because of the TSO's reliability management, every line investment will both increase

[2]The European Union wants to bring the electricity interconnection level of all member countries to 10% by 2020 and is looking into raising the target to 15% by 2030 (European Commission, 2015).

reliability and decrease congestion. Also Kirschen and Strbac (2004b) discuss the effect of investment on congestion and reliability. They observe that the N-1 reliability criterion treats congestion and reliability differently. We will show this in a more formal way. The N-1 criterion is also criticized by Joskow (2006), who finds that little effort has been made to review it and evaluate its costs and benefits. An exception is the case study of de Nooij et al. (2010), which compares for the N-1 rule the transmission investment costs with the benefits of reduced interruption costs. Lastly, Joskow and Tirole (2007) formulate a model similar to the one of this paper, but they focus on generation adequacy (Steiner, 1957; Williamson, 1966) instead of transmission reliability. In their model, the reliability level is the probability that demand is lower than *installed* generation capacity (Chao, 1983; Kleindorfer and Fernando, 1993), while in this paper the reliability level is the probability that the electricity flow is lower than *available* transmission capacity.

Reliability criteria are also increasingly being studied in the engineering literature (Grigg et al., 1999; Kirschen and Jayaweera, 2007; He et al., 2010; Capitanescu et al., 2011b; Heylen et al., 2016b, 2017b). These case studies all study the short-term effect of alternative reliability criteria, without making a link to transmission investment. By modeling the network in great detail, they arrive at estimates of possible short-term efficiency improvements. This paper intends to complement these studies by focusing on a stylized setup instead of on simulation results, which allows the clarification of the driving forces and trade-offs, both in the short and long term.

The paper is structured as follows. Section I.2 introduces the model that shows the trade-off between congestion cost, transmission investment cost and reliability. We determine the optimal investment and optimal use of transmission capacity. Section I.3 analyses the expected interruption cost function, the optimal price difference between zones and cost recovery in the optimal solution. Section I.3.4 extends the basic model to multiple states of the world and a general network. Next, section I.4 analyses the N-1 reliability criterion and shows that its reliability margin does not depend on economic and technical parameters. In section I.5, we illustrate the analysis with a numerical example in a simple and in a more complex network.

Section I.6 describes the requirements to move beyond the N-1 reliability criterion. Finally, section I.7 concludes.

I.2 The model

I.2.1 The model setup

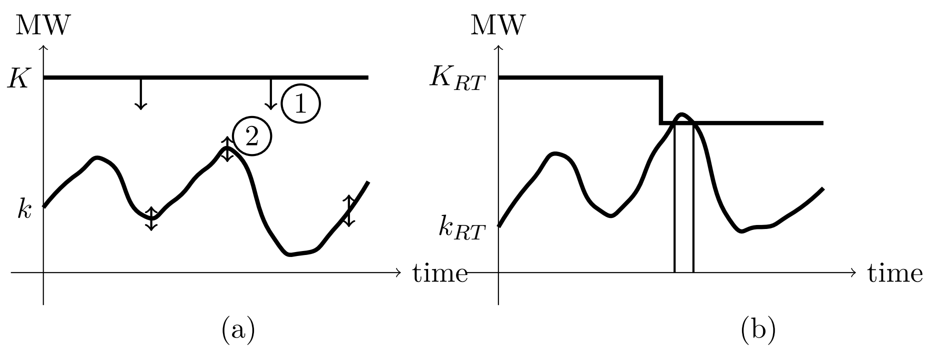
Consider two regions connected by multiple transmission lines. We define the maximum capacity K [MW] as the maximum possible electricity flow between the two regions when all electricity lines are in operation. Some time before real time (for example in the day-ahead market for the 24 hours of the next day), the TSO determines how much electricity flow k [MW] to schedule between the two regions.[3] The TSO always schedules less flow than the maximum capacity K to account for (*i*) unplanned outages of transmission lines and transformers – which decrease the maximum capacity – and to account for (*ii*) forecast errors, loop flows and unplanned outages of generation plants and large loads – which causes the physical flow to differ from the scheduled flow.

The left-hand panel of Figure I.1 shows the maximum capacity K and the scheduled flow k. The reliability margin is defined as the difference between maximum capacity and scheduled flow: $K - k$ (Neuhoff et al., 2013). The TSO keeps this margin because the physical flow could differ from the scheduled flow and the maximum capacity could be lower than expected due to transmission line failures. The right-hand panel of Figure I.1 shows a possible real-time realization of maximum capacity and physical flow. In this case a combination of a higher physical flow and a line failure causes the physical flow k_{RT} to be larger than the real-time maximum capacity K_{RT} at some point. As the transmission capacity of the remaining lines is insufficient to accommodate the physical flow between the regions, the TSO needs to do a corrective action such that the physical flow is back within the bounds of the real-time maximum capacity: $k_{RT} < K_{RT}$. Possible corrective actions

[3] If the two regions are in a different price zone, the TSO directly decides on the scheduled flow by imposing a limit on cross-regional flows. If the two regions are within a price zone, the TSO only indirectly decides on the scheduled flow by using preventive redispatch of generation.

are use of generation reserves or involuntary load shedding – also called non-price rationing, demand curtailment or controlled rolling blackouts (Joskow, 2008b). Generation reserves allow increasing or decreasing generation in different parts of the network, while load shedding amounts to deliberately restricting electricity supply in parts of the network. If the physical overflow is not adequately dealt with within a certain period of time, this could lead to cascading uncontrolled network collapses and large-scale blackouts (e.g. in the U.S. (2003) , Italy (2003), Brazil and Paraguay (2009), India (2012) and Turkey (2015)).

Figure I.1: (a) Maximum capacity K and scheduled flow k before real time. (b) Real-time maximum capacity and physical flow



This paper studies the optimal investment (K^*) and optimal scheduled use (k^*) of transmission capacity between two regions. Suppose that these decisions are made by a welfare-maximizing TSO, the entity responsible for dispatch, congestion management, maintenance and investment of the transmission network.[4] Optimally a TSO schedules less electricity flow k than the maximum capacity K. It keeps a scheduled reliability margin $K-k$ or equivalently, it keeps a line loading $\alpha = \frac{k}{K} < 1$. Before determining the optimal investment and optimal use of transmission capacity, we study the three types of costs and benefits that constitute net interconnection surplus.

[4]If the operational, maintenance and investment responsibility tasks are split, the respective entities are called the Independent System Operator (ISO) and the independent transmission company (Transco). The ownership or division of tasks does not fundamentally affect the core of our analysis.

I.2.2 The three constituents of net interconnection surplus

First, scheduling an electricity flow k from a low-cost region to a high-cost region creates interconnection benefit, or gross interconnection surplus, by enabling a reduction of production costs. The difference of production costs[5] between a system with some constrained transmission lines and one with infinite transmission capacity, plus the consumer dead-weight loss from the associated changed prices, are called the congestion costs of the system (Joskow, 2006).[6] In addition to decreased congestion costs, interconnection decreases the cost of reserves through reserve sharing and demand smoothing (Baldursson et al., 2016b), it creates more competition in the generation market (Borenstein et al., 2000) and facilitates the integration of renewable generation. The function $S(k)$ summarizes the interconnection benefit. It is an increasing and concave function:

$$S'(k) \geq 0 \text{ and } S''(k) < 0 \tag{1.1}$$

Second, transmission line failures and physical flows that diverge from scheduled flows could require corrective actions by the TSO. In the remainder of this paper we will assume that load shedding is the only corrective action available. In addition, we assume that the TSO is able to estimate the expected interruption cost (EIC) of load shedding. The EIC depends on both scheduled electricity flow k and maximum capacity K. Ceteris paribus, the EIC is increasing and convex in k and decreasing and convex in K.

$$\begin{aligned} EIC'_k(k,K) > 0 \text{ and } EIC''_k(k,K) > 0 \\ EIC'_K(k,K) < 0 \text{ and } EIC''_K(k,K) > 0 \end{aligned} \tag{1.2}$$

[5] Assuming perfectly competitive producers and assuming that the generation cost is the social cost of generation, i.e. including externalities like CO_2, NO_X, SO_X, particulate matter, noise, etc.

[6] To manage congestion, cheap generation in an export-constrained region is decreased, while more expensive generation in an import-constrained region is increased. In uniform-price zones, redispatch is the responsibility of the TSO, who gives a congestion payment to redispatched generators. In nodal pricing or market splitting, congestion is managed implicitly by prices. Different congestion management methods lead to different physical and financial flows, and have different effects on TSOs (Holmberg and Lazarczyk, 2015) and on producers (Dijk and Willems, 2011).

Third, installing and using transmission capacity has a cost. Suppose that the transmission costs only depend on the maximum capacity K. That is, we incorporate operations, maintenance and investment costs but neglect losses and line-loading induced depreciation and maintenance, which depend on the ratio of k and K. The cost of transmission capacity, $c(K)$, is increasing in K:

$$c'(K) > 0 \tag{1.3}$$

I.2.3 The optimal solution

Following the above assumptions, net surplus of transmission interconnection is given by:[7]

$$\max_{\{k,K\}} \{S(k) - EIC(k,K) - c(K)\} \tag{1.4}$$

Net surplus is interconnection benefit minus expected interruption costs and transmission investment costs. It is maximized by selecting the optimal scheduled flow k and maximum capacity K.[8] These are calculated from the first-order conditions of net surplus (1.4):

$$\begin{cases} S'(k) = EIC'_k(k,K) \\ -EIC'_K(k,K) = c'(K) \end{cases} \tag{1.5}$$

The first trade-off between interconnection benefit and expected interruption costs is the TSO's short-term decision. Given a constant maximum capacity K, electricity flow k is optimally scheduled up to the point where the additional increase of interconnection benefit equals the additional increase of expected interruption costs. The short-term first-order condition is thus a trade-off between congestion and reliability: increasing the scheduled

[7] The constraint $k \leq K$ is never binding in expected terms. That is, it is assumed that at $k = K$, marginal interconnection benefit is lower than marginal expected interruption costs. However, k_{RT} can be higher than K_{RT} in real time, as was illustrated in Figure I.1.

[8] Note that equation (1.4) represents the maximization of net interconnection surplus. If the interconnection is between two TSO zones z, the optimal electricity flow k_z (and thus K_z) can differ from the point of view of each TSO, depending on the division of interconnection benefit and expected interruption costs. This may lead to strategic TSO behavior (Glachant and Pignon, 2005).

flow decreases congestion but increases the expected interruption cost. As a consequence, complete elimination of congestion costs at all times entails a too large expected interruption cost, while aiming for a 100 % reliable transmission system entails a too large congestion cost.

Both $S(k)$ and $EIC(k, K)$ could change over time. For example, $S(k)$ increases if autarkic prices are more apart and $EIC(k, K)$ increases in case of adverse weather (Kirschen and Jayaweera, 2007).

The trade-off between expected interruption costs and transmission investment costs determines the TSO's long-term decision of how large the maximum capacity K should be. Since the two first-order conditions are interdependent, the optimal maximum capacity is determined by a trade-off between reliability, congestion and investment. Increasing maximum capacity and keeping the scheduled electricity flow constant, decreases expected interruption costs. Increasing maximum capacity and keeping expected interruption costs constant, decreases congestion costs.

The first-order conditions (1.5) lead to the following result:

Proposition 1. *Consider two regions between which a flow K can be transferred. Optimally, the TSO should:*

i. In the short term, schedule electricity flow until the marginal interconnection benefit equals the marginal expected interruption costs.

ii. In the long term, increase maximum capacity K until the marginal expected interruption cost equals the marginal cost of interconnection.

I.3 Analysis

I.3.1 The expected interruption cost

To make the above optimal solution more concrete, we calculate the expected interruption cost of a simple network, as a function of k and K.[9] Assume two regions connected by n identical transmission lines with a joint maximum capacity of K MW. Each line has a transmission capacity of K/n MW, the

[9] For a numerical illustration see (Kirschen and Jayaweera, 2007). They calculate the expected interruption cost of a IEEE 24-bus reliability test system.

same technical characteristics and the same probabilistically-independent failure probability p_f. Suppose that the scheduled flow k equals the physical flow in real time. That is, only the availability of maximum capacity is uncertain. When, due to line failures, only i of the n lines are available, the TSO needs to shed $\left(k - \frac{i}{n}K\right)^+$ MW of load, assuming that load shedding is the only available corrective action. For example, if maximum capacity is 4,000 MW, scheduled flow is 3,000 MW, and only 3 of 5 identical lines are available, $3000 - \frac{3}{5}4000 = 600$ MW of load shedding is needed to keep physical flow below the available transmission capacity of 2,400 MW. The cost of shedding a MW of load is represented by the value of lost load (VOLL), which is the lost surplus when a MWh of energy is not served to consumers demanding this energy.[10] VOLL (V) is generally expressed in €/MWh. The above assumptions lead to the following EIC:

$$EIC(k, K) = \sum_{i}^{n} p_i \left(k - \frac{i}{n}K\right)^+ V \tag{1.6}$$

where $p_i = \binom{n}{i}(1 - p_f)^i (p_f)^{(n-i)}$ is the probability that i of n lines are available, given i.i.d. failures. This specific expected interruption cost functional form fulfills the above assumptions (1.2): it is increasing in k, decreasing in K and convex piecewise-linear in k and K.

The short-term first-order condition showed that the optimal scheduled flow k is at the point where the marginal interconnection benefit equals the marginal EIC. Assuming the EIC of equation (1.6) this is:

$$\begin{aligned} S'(k) &= \sum_{i}^{n} p_i \mathbb{1}_+(i) V \\ \text{where } \mathbb{1}_+(i) &= \begin{cases} 1 & \text{if } i < n\frac{k}{K} \\ 0 & \text{if } i \geq n\frac{k}{K} \end{cases} \end{aligned} \tag{1.7}$$

where $\mathbb{1}_+$ is the indicator function, which is equal to one if load shedding is needed, i.e. if $(k - \frac{i}{n}K)$ is positive.

Equation (1.7) shows that the optimal electricity flow depends on the network topology (n), the line failure probability (p_f), the value of lost load

[10] A rich literature exists on measuring the value of supply interruptions using stated preference (Ovaere et al., 2016; Pepermans, 2011; Reichl et al., 2013), revealed preference, indirect analytical methods (de Nooij et al., 2007) or case studies.

(V), the marginal interconnection benefit (S'), and maximum capacity (K). For example, in case of adverse weather (p_f high) or when consumers have a high VOLL, the TSO should schedule a lower electricity flow. To summarize:

Proposition 2. *When the EIC is as in equation* (1.6)*, the TSO should optimally increase the electricity flow when failure probabilities decrease, marginal interconnection benefit increases, VOLL decreases, and maximum capacity increases.*

I.3.2 The optimal price difference and cost recovery of transmission investments

In the standard economic transmission investment model with 100% reliable transmission capacity ($K = k$), as in (Kirschen and Strbac, 2004a, p.152) and (Turvey, 2006), the optimal investment in transmission capacity is at the point where the marginal benefit of interconnection equals the marginal cost of transmission:

$$S'(K^*) = c'(K^*) \tag{1.8}$$

Only incorporating congestion costs into the interconnection benefit, the marginal benefit is equal to the price difference $\Delta p(k)$ between the two zones. Additionally, assuming a constant marginal cost of interconnection c, the expression becomes:

$$\Delta p(K^*) = c \tag{1.9}$$

This standard result is altered when reliability is included. Assuming the expected interruption cost of equation (1.6), we can combine the two first-order conditions to get the following expression:

$$\Delta p(k^*) = c + \sum_{i}^{n} p_i \mathbb{1}_+(i)(1 - \frac{i}{n})V \tag{1.10}$$

This shows that reliability concerns cause the optimal price difference to be larger than the marginal investment cost. The wedge between the price difference and the marginal investment cost increases with VOLL and with failure probability.

Cost recovery of transmission investments also changes by including reliability concerns. Without reliability concerns, it follows from equation

(1.9) that congestion rent $K\Delta p$ equals variable transmission investment cost cK^*.[11] With reliability concerns and assuming that EIC is a homogeneous function of degree h, the first-order conditions (1.5) combine to:[12]

$$k^* \Delta p(k^*) = cK^* + hEIC(k^*, K^*) \tag{1.11}$$

That is, optimally, the congestion rent earned by the TSO (left-hand side) is larger than the cost of transmission investment (cK^*). For example, if the expected interruption cost is the one of equation (1.6), which is homogeneous of degree $h = 1$[13], the optimal congestion rent equals the sum of investment cost and interruption costs. This excess revenue could be used to compensate consumers for their interruption costs. To summarize:

Proposition 3. *If reliability is considered, the optimal price difference leads to more than sufficient congestion rent to remunerate the cost of transmission capacity. If the expected interruption cost is homogeneous of degree 1, the excess congestion rent can fully compensate consumers for their interruption costs.*

I.3.3 Economic versus reliability transmission investments

Some TSOs and regulators consider reliability transmission investments and economic transmission investment as being two separate objectives: FERC (2006), ENTSO-E (2014a, p.60), PJM (Joskow, 2005, p.111). Economic transmission investments are conceptualized as being developed to reduce congestion costs, while reliability transmission investments are conceptualized as necessary to meet engineering reliability criteria. However, first-order conditions (1.5) show that a categorization into reliability transmission investment and economic transmission investment is arbitrary. Investing in

[11] In reality, congestion rent falls considerably short of variable investment costs, as illustrated by Pérez-Arriaga et al. (1995), due to lumpiness and decreasing marginal transmission investment costs. In addition, fixed transmission investment costs are not recovered. Also note that the TSO only earns congestion rent on transmission lines between different-price zones.

[12] Using Euler's Homogeneous Function Theorem: $EIC'_k k + EIC'_K K = hEIC$

[13] This means that with equal line loading α, an optimal transmission network that is twice as large has twice as much expected optimal interruption costs. That is, there is constant returns to scale in interruption costs.

more transmission capacity can lead to more interconnection benefit and to a lower expected interruption cost, depending on the TSO's choice of scheduled flow k. In the short term, one can increase reliability by increasing congestion, and vice versa.

I.3.4 General model formulation

The analysis up to now was restricted to a two-region network with constant interconnection benefit and EIC. This section generalizes our results to a stochastic model and to general networks.

I.3.4.1 Stochastic model

We distinguish different states of the world. Every state of the world is a particular combination of parameter values for demand and supply conditions, failure probabilities, VOLL, etc. If $t \in T$ is the state of the world, the objective function becomes:

$$\max_{\{k_t, K\}} \{S_t(k_t) - EIC_t(k_t, K) - c(K)\} \tag{1.12}$$

This means that for each state of the world t one should schedule electricity k until the marginal surplus equals the marginal expected interruption cost:

$$\begin{cases} S_t'(k_t) = EIC_{t,k}'(k_t, K) \qquad \forall t \in T \\ -\int_T EIC_K'(k_t, K) f(t) dt = c'(K) \end{cases} \tag{1.13}$$

The first line of (1.13) shows that k_t is different for different states of the world. For example, when interconnection benefit is high, increase scheduled flow; and when expected interruption costs are high, decrease it. The first order condition for investment (second line of (1.13)) shows that one should increase maximum capacity K until the marginal cost of interconnection equals the marginal expected (over all states of the world) interruption costs.

I.3.4.2 General network

If the network consists of N nodes and L lines, the objective function becomes

$$\max_{\{\vec{k},\vec{K}\}} \{S(\vec{k}) - EIC(\vec{k},\vec{K}) - c(\vec{K})\} \tag{1.14}$$

where $\vec{k} \in \mathbb{R}^L$ is the vector representing the flows over the L lines of the network and $\vec{K} \in \mathbb{R}^L$ is the vector representing the transmission capacity of the L lines of the network. However, the TSO does not directly control the flows on the transmission lines $\vec{k}$. During day-ahead generation dispatch the TSO decides on a generation schedule $\vec{g} \in \mathbb{R}^N$, which leads to a unique power flow schedule $\vec{k}$.[14] Therefore, reformulate the objective function as:

$$\max_{\{\vec{g},\vec{K}\}} \{S(\vec{g}) - EIC(\vec{g},\vec{K}) - c(\vec{K})\} \tag{1.15}$$

The first-order conditions are:

$$\begin{cases} S'_{g_n}(\vec{g}) = EIC'_{g_n}(\vec{g},\vec{K}) & \forall n \in N \\ -EIC'_{K_l}(\vec{g},\vec{K}) = c'_{K_l}(K_l) & \forall l \in L \end{cases} \tag{1.16}$$

That is, for each node n schedule generation capacity g_n until the marginal surplus equals the marginal expected interruption cost. In the optimum, congestion-increasing generation ($S'_{g_n} < 0$) decreases EIC; congestion-decreasing generation ($S'_{g_n} > 0$) increases EIC. The long-term first-order condition shows that for each line l one should increase transmission capacity until the marginal cost of interconnection equals the marginal expected interruption costs.[15]

I.4 The N-1 reliability criterion

Currently all TSOs use the N-1 reliability criterion or some variant.[16] The N-1 criterion states that an unexpected outage of a single system component

[14] In the two-region analysis with inelastic demand D, generation g translates directly into a flow k: $g - D = k$.

[15] In reality, both the large number of states of the world and lumpy investments with large fixed costs make it impossible to satisfy the long-term first-order condition for all lines l. Discrete investment decisions are studied in more detail in section I.5.2.

[16] For example: N-0 during maintenance, considering double-line failures during adverse weather, N-1 of credible contingencies (like multiple dependent failures), stronger reliabil-

(lines, transformers, generation plants, large loads, etc.) may not result in a loss of load. That is, when a single system component fails, the transmission system should still be able to accommodate all flows without load shedding. This also implies that, following the N-1 reliability criterion, a simultaneous failure of multiple system components could require load shedding to avoid a black out.

I.4.1 Operational planning decision

First we analyse how the N-1 reliability criterion determines the TSO's operational planning decision of how much electricity flow to schedule. By prohibiting lost load in case of a single contingency, the N-1 reliability criterion determines the electricity flow k_{N-1} allowed on the network as the maximum flow that the network can accommodate after each possible single contingency. We represent this decision as the allowed maximum line loading $\alpha_{N-1} = k_{N-1}/K$, since in the short-term maximum capacity K is constant. Figure I.2 shows how to determine the allowed line loading α_{N-1} for two different networks, with equal maximum capacity between two regions East and West.

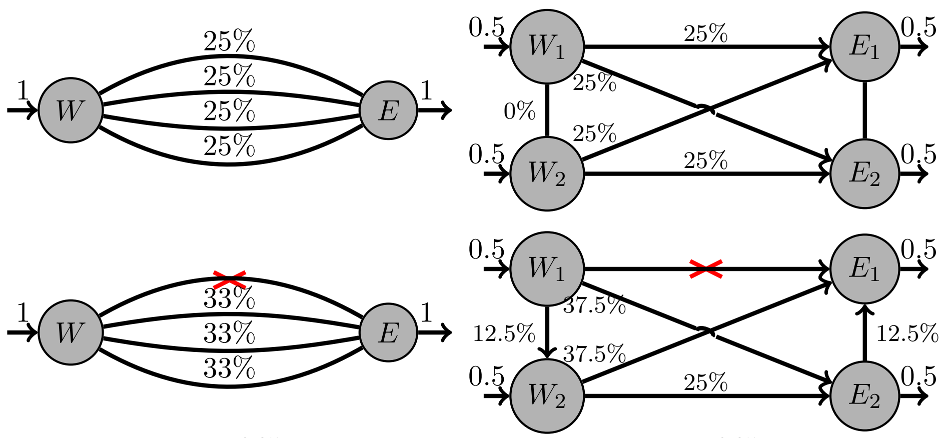
Network (a) shows two nodes connected by four identical transmission lines. Each line carries 25% of the total electricity flow. The line loading implied by the N-1 reliability criterion is $\alpha_{N-1} = \frac{0.25}{0.33} = 0.75$, since the network should be able to cope with a loss of one line, or a quarter of the maximum capacity. For example, if each line has a thermal limit of 1,000 MW, 3,000 MW is available in each N-1 state.

Network (b) shows the same four identical lines, but with two additional identical lines within East and West. Demand and supply are split evenly within each zone. Nodes W_1 and W_2 each supply half of net exports, while nodes E_1 and E_2 each consume half of net imports. If a transmission line in this interconnected system fails, the power flows are redistributed automatically throughout the network according to Kirchoff's laws. A lossless DC Power flow analysis[17] calculates that in this network the maximum flow

ity criteria for cities or certain business districts, etc. For example, the Dutch regulator has changed the reliability criterion to 'N-1 during maintenance, unless the costs exceed the benefits' (de Nooij et al., 2010).

[17] A DC power flow is a linear approximation of the Kirchoff's laws that assumes that (i)

Figure I.2: Different allowed line loadings α_{N-1}, depending on the topology of the network.



(a) $\alpha_{N-1} = \frac{0.25}{0.33} = 0.75$ (b) $\alpha_{N-1} = \frac{0.25}{0.375} = 0.67$

on a single line in case of a line failure is 37.5% of the total flow, which occurs when one of the interzonal lines fails, as shown for line $W_1 - E_1$ in the lower part of Figure I.2 (b).[18] As a result, the N-1 reliability criterion dictates a line loading of $\alpha_{N-1} = \frac{0.25}{0.375} = 0.67$, which is the ratio of the maximum flow of the N state and the maximum flow of all of the N-1 states. Note that in meshed networks the line loading parameter represents capacity utilization, i.e. the percentage of maximum capacity that is used to schedule power flow. The flow on individual lines differs from this value. In this network topology, if capacity utilization is below 67% of maximum capacity, the flow on each of the six transmission lines never exceeds its thermal limit in case one of the six lines fails. For example, if each line has a thermal limit of 1,000 MW, the scheduled electricity flow allowed by the N-1 reliability criterion is 2667 MW. In case one of the interzonal lines fails, two interzonal lines are at their thermal limit of 1000 MW, the third interzonal line is at 667 MW and the intrazonal lines are at 333 MW.

This example shows that the allowed N-1 line loading or capacity uti-

voltage angle differences between neighbouring nodes are small, (*ii*) the voltage is equal for all nodes, and (*iii*) line resistances are negligible compared to line reactances (Van den Bergh et al., 2014).

[18]The appendix shows the manual power flow calculation for this simple network.

lization α_{N-1} depends on network topology[19], but not on the probability of line failures or the cost of interruptions.

Proposition 4. *The N-1 reliability criterion determines a maximum line loading α_{N-1} that depends on the network topology, but does not depend on the probability of line failures, the interconnection benefit, or the cost of interruptions.*

I.4.2 Investment decision

A TSO strictly following the N-1 reliability criterion ensures that in every state of the world the criterion is met. This means that the maximum line loading is α_{N-1}, irrespective of the TSO's investment decision. As a consequence, the TSO does not directly assess the effect of transmission investments on reliability. While the optimal investment decision is a trade-off between interconnection benefit, reliability and investment cost (first-order conditions (1.5)), the N-1 investment decision is a trade-off between only interconnection benefit and investment cost:

$$\max_{\{K_{N-1}\}} \{S(\alpha_{N-1}K_{N-1}) - cK_{N-1}\} \longrightarrow S'(\alpha_{N-1}K_{N-1}) = c \tag{1.17}$$

That is, an N-1 investment aims at alleviating congestion, not improving reliability, which is exogenously-determined by the N-1 reliability criterion. The investment could increase reliability, however, but it is not its objective.

An intermediate step between N-1 reliability management and optimal reliability management is when the TSO determines the line loading using the optimal short-term first-order condition, but the maximum capacity using equation (1.17):

$$\max_{\{K_{ST}^*\}} \{S(\alpha^* K_{ST}^*) - cK_{ST}^*\} \longrightarrow S'(\alpha^* K_{ST}^*) = c \tag{1.18}$$

That is, by considering reliability optimally in the short term, but not in the long term. As interconnection benefit S is increasing concave in α, comparison of equations (1.17) and (1.18) shows that K_{ST}^* is lower than K_{N-1} if α^* is higher than α_{N-1}, and vice versa.

[19] The N-1 line loading also depends on technical characteristics of the transmission lines and the spatial distribution of demand.

Comparing optimal maximum capacity K^* with the above maximum capacities is less straightforward in general terms, as the EIC function can be very complex. However, comparison is possible by assuming that EIC is homogeneous of degree h. Equation (1.11) can be written as:

$$S'(\alpha^* K^*) = \frac{c}{\alpha^*} + \frac{hEIC}{\alpha^* K^*} > c \tag{1.19}$$

As the right-hand side of equation (1.19) is higher than the marginal cost of capacity investment c, comparing equation (1.19) with the first-order condition of equation (1.18) shows that maximum capacity is always lower when reliability management is optimal in the short- and the long-term than when it is only optimal in the short-term: $K^* < K^*_{ST}$. That is, if a TSO makes the optimal trade-off for operational decisions but not for investment decisions, there is always an overinvestment of transmission capacity. On the other hand, comparing equation (1.19) with the first-order condition of equation (1.17) reveals that optimal maximum capacity is lower than N-1 maximum capacity if the optimal line loading is higher than the N-1 line loading: $K^* < K_{N-1}$ if $\alpha^* > \alpha_{N-1}$. If optimal line loading is lower than the N-1 line loading, the result is ambiguous. However, as a low optimal line loading also increases the right-hand side of equation (1.19), N-1 maximum capacity is only lower than optimal maximum capacity if optimal line loading is considerably lower than N-1 line loading. To summarize, the N-1 criterion leads to overinvestment, except when the optimal line loading is considerably lower than N-1 line loading.

Proposition 5. *Suppose the expected interruption cost function is homogeneous. If the N-1 criterion is too conservative ($\alpha_{N-1} < \alpha^*$), it leads to overinvestment in transmission capacity: $K^* < K_{N-1}$.*

I.5 Numerical illustration

I.5.1 A simple network

To illustrate the analysis, we assume functional forms for the three components of net interconnection surplus: interconnection benefit, expected interruption cost and transmission investment cost.

First, suppose that interconnection benefit $S(k)$ consists only of the decrease of congestion costs. Competition effects are neglected, and the importing country is not structurally depending on imports, meaning that decreasing the scheduled electricity flow does not lead to preventive load shedding in the importing country. To express the interconnection benefit, suppose that the slope of the residual supply curve S^E_{res} in the exporting country is b_E and the slope of the residual demand curve D^I_{res} in the importing country is b_I. Figure I.3 shows how the supply and demand curves of the importing and exporting region lead to the residual supply curve and the residual demand curve. Scheduling an electricity flow k causes prices to converge, leading to increased producer surplus (A+B) and decreased consumer surplus (A) in the exporting region and increased consumer surplus (D+E) and decreased producer surplus (E) in the exporting region (Turvey, 2006). The sum of the two CR areas is the congestion rent, which is the product of the price difference and the scheduled flow. The interconnection benefit $S(k)$ is the sum of additional producer surplus (B), congestion rent (CR) and additional consumer surplus (D):

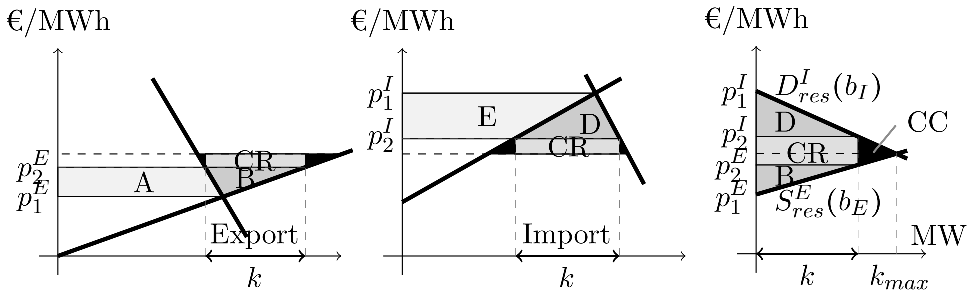
$$\begin{aligned} S(k) &= \Delta PS_E + CR + \Delta CS_I \\ &= \frac{1}{2}(p_2^E - p_1^E)k + k((p_1^I - p_1^E) - (b_E + b_I)k) + \frac{1}{2}(p_1^I - p_2^I)k \\ &= \frac{k}{2}b_E k + k(\Delta p - (b_E + b_I)k) + \frac{k}{2}b_I k \\ &= k(\Delta p - 0.5(b_E + b_I)k) \end{aligned} \tag{1.20}$$

where $\Delta p = p_1^I - p_1^E$ is the initial price difference before interconnection. The above expression shows that interconnection benefit is increasing with Δp, decreasing with $b_E + b_I$ and concave in k (which is in line with the assumption of equation (1.1)). The scheduled electricity flow that completely eliminates congestion cost (CC) by equalizing prices in the exporting and importing region is $k_{max} = \Delta p/(b_E + b_I)$.

Second, assume the following expected interruption cost function of two regions connected by n identical transmission lines:

$$EIC(k, K) = \sum_i^n p_i \left[(k - \frac{i}{n}K)^+ V + \left(\frac{(k - \frac{i}{n}K)^+}{A} \right)^a V_{BO} \right] \tag{1.21}$$

Figure I.3: Change of interconnection benefit with scheduled flow k: consumer surplus (CS), producer surplus (PS), congestion rent (CR) and congestion cost (CC).



where p_i is the probability that i lines are available, V the value of lost load and V_{BO} the cost of a blackout. This EIC is increasing in k, decreasing in K and convex in k and K (which is in line with the assumption of equation (1.2)). The first part of the EIC is the expected cost due to load shedding (equation (1.6)), the second part is the expected cost of a black out. We add this expected cost of a black out because each overflow that requires corrective load shedding has a probability of leading to a blackout, for example if corrective load shedding fails or if overloaded lines trip before load shedding is executed. We assume that an overflow $(k - \frac{i}{n}K)^+$ has a probability $\left((\frac{k-\frac{i}{n}K}{A})^+\right)^a$ to lead to a widespread blackout. The parameter A is a value that is larger than the maximum scheduled flow k, such that the probability is always lower than one. The higher this parameter A, the lower the probability that overflows lead to a blackout. In addition, $a > 1$, such that the probability is increasing with increasing overflow. That is, larger overflows have an increasingly higher probability of leading to a blackout. Suppose that V_{BO} [€/h] is the cost of a blackout.[20]

Third, suppose that increasing maximum transmission capacity K has a constant marginal cost c.[21] Following the above functional form assump-

[20] As an illustration, values can be found on `http://www.blackout-simulator.com/` (Reichl et al., 2013).

[21] Obviously there are economies of scale in transmission. A cost function of $c(K) = F + cK$ has increasing returns to scale and would yield qualitatively the same results as

tions, net interconnection surplus is:

$$\max_{\{k,K\}} \{k(\Delta p - 0.5(b_E + b_I)k) - \sum_{i}^{n} p_i \left[(k - \frac{i}{n}K)^+ V + \left(\frac{(k - \frac{i}{n}K)^+}{A} \right)^a V_{BO} \right] - cK\} \tag{1.22}$$

We assume the illustrative parameter values of Table I.1 for respectively interconnection benefit, EIC and investment cost. Numerical optimization of this objective function leads to the following optimal capacities:

$$k^* = 1667 \text{ MW and } K^* = 2089 \text{ MW}$$

The optimal line loading α^* equals 0.7979, compared to the N-1 line loading $\alpha_{N-1} = 0.75$. The specific four-line topology and the chosen parameter values result in an optimal line loading that is less conservative than the N-1 line loading. However, numerical results for optimal capacities differ greatly with the parameter values. Figure I.4 shows the optimal short-term line loading (bold curve) for different values of the cost of blackout V_{BO}, line failure probability p_f, initial price difference Δp, and the sum of the slope of the residual supply and demand curves $b_E + b_W$. Maximum capacity is fixed at the optimal capacity K^* for the parameter values of Table I.1. The dot in the four graphs indicates the optimal line loading $\alpha^* = 0.7979$, while the dashed line represents the N-1 line loading $\alpha_{N-1} = 0.75$. The graphs of Figure I.4 show that the optimal line loading changes if the economic and technical parameters change. The top left-hand graph shows that the optimal line loading decreases with the cost of a blackout. The top right-hand graph shows that the line loading should optimally be higher than the N-1

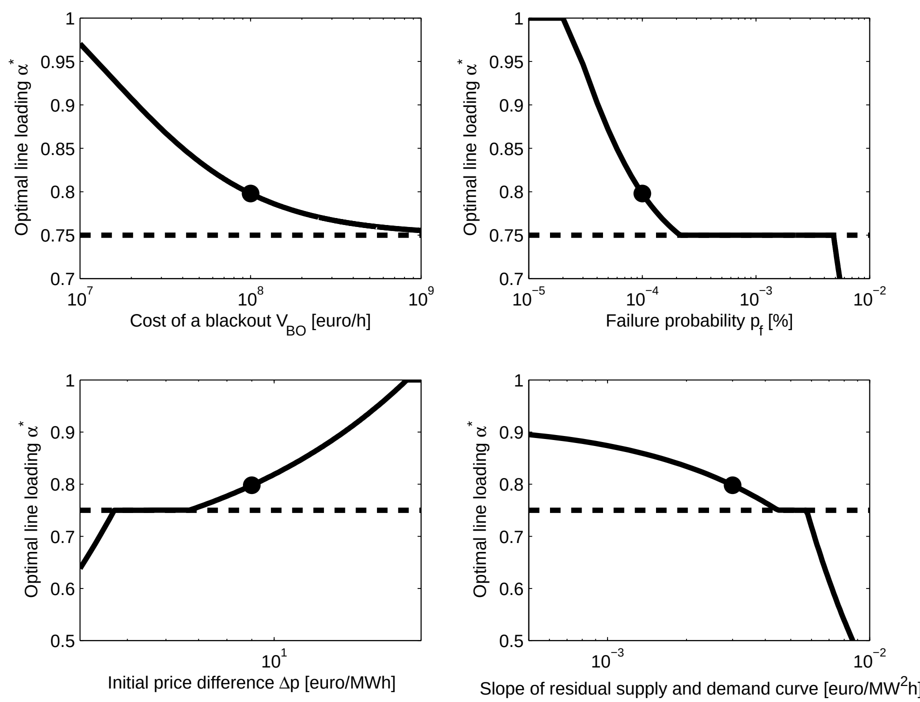
the current analysis.

Table I.1: Illustrative parameter values

$\Delta p = 9$	[€/MWh]	$V = 5000$	[€/MWh]	$c = 3$	[€/MWh]
$b_E + b_I = 0.003$	[€/MW^2h]	$p_f = 10^{-4}$	[/]		
		$V_{BO} = 10^8$	[€/h]		
		$n = 4$	[/]		
		$a = 2$	[/]		
		$A = 2000$	[/]		

line loading when the probability of line failure is low, while the line loading should optimally be lower than the N-1 line loading if the probability of line failure is high. For intermediate failure probability values, the N-1 line loading is optimal in this simple two-node example. The same is true for the initial price difference; if Δp is higher, the line loading should optimally be higher.

Figure I.4: Optimal short-term line loading α^* for different values of V_{BO}, p_f, c and $b_E + b_W$.



Maximum capacity also depends on the above parameter values. While the short-term optimal line loading changes in real time in response to changing parameter values, such as the actual price difference or the line failure probabilities, maximum capacity changes in response to persistent changes of parameter values, such as the cost of a blackout or the average price difference. Figure I.5 shows the optimal maximum capacity K^* (from equation (1.22)), optimal short-term maximum capacity K^*_{ST} (from equation (1.18)), and N-1 maximum capacity K_{N-1} (from equation (1.17)) for different values of the parameters. The optimal short-term maximum capacity K^*_{ST} (dashed line) is close to the optimal maximum capacity K^* (solid line) if the optimal line loading differs substantially from the N-1 line loading. For our parameter values this is the case when $b_E + b_I$ is high and V_{VBO}, p_f and Δp are low. By contrast, if the optimal line loading is close to the N-1 line loading, optimal short-term maximum capacity K^*_{ST} is closer to the N-1 maximum capacity (dotted line). Note that Figure I.5 confirms the results of section I.4.2 and of proposition 5. Optimal maximum capacity K^* is always lower than optimal short-term maximum capacity K^*_{ST}, while the relative value of K^*_{ST} and K_{N-1} depends on the ratio of line loadings.

Figure I.5: Optimal, optimal short-term and N-1 maximum capacity for different values of V_{BO}, p_f, c and $b_E + b_W$.

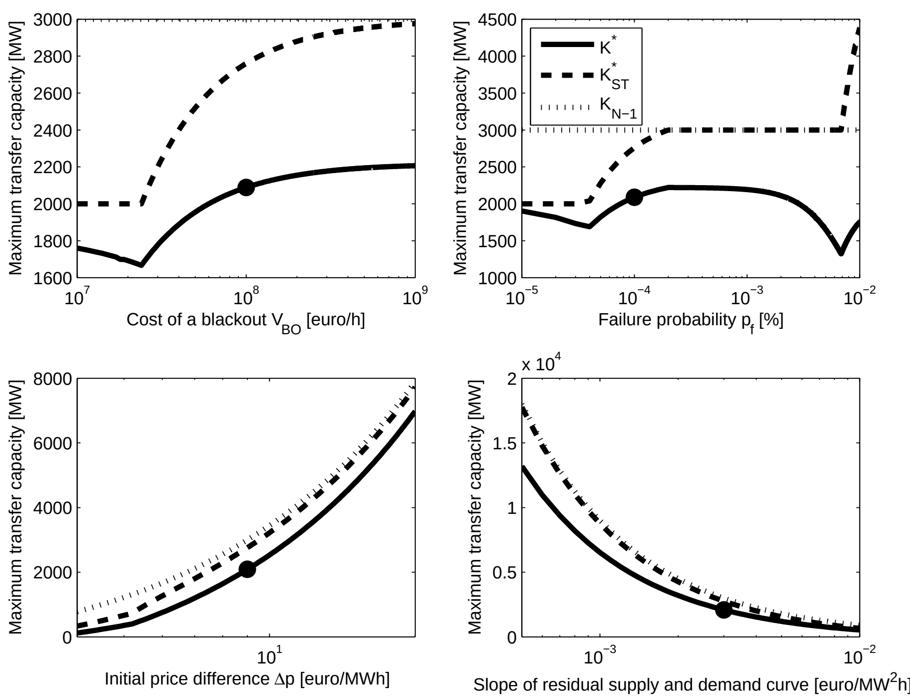


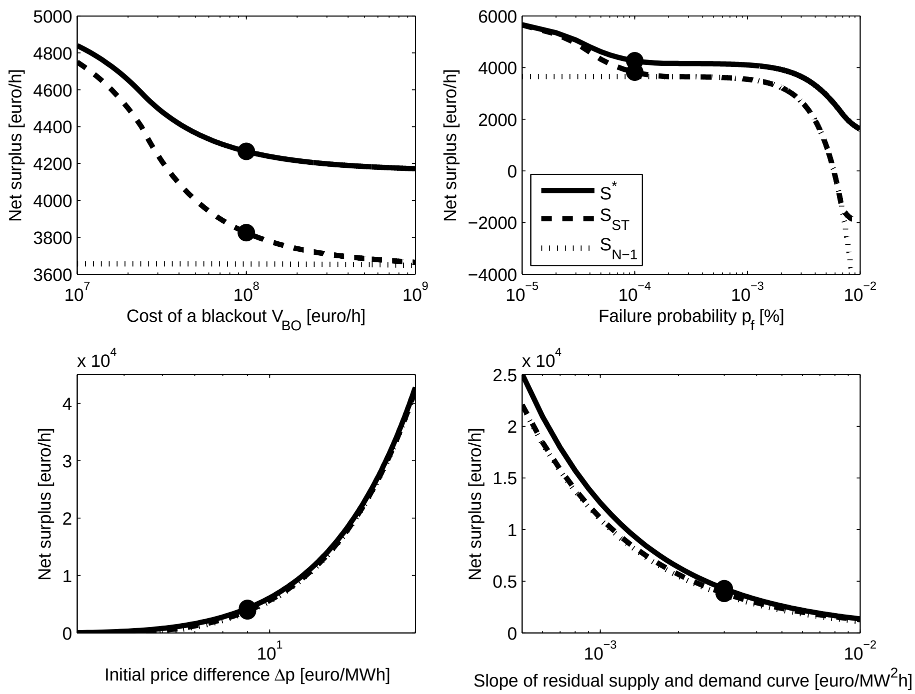
Figure I.6 shows the effect of different degrees of optimal reliability decisions on long-term surplus, for different values of the parameters. This is the value of equation (1.22) for optimal reliability S^* (solid line), for optimal short-term reliability but non-optimal investment S^*_{ST} (dashed line) and for N-1 reliability in the short and long term S_{N-1}. These graphs show that S^*_{ST} is weakly lower than S^*, and S_{N-1} is weakly lower than S^*_{ST}. For the parameter values of table I.1, surplus increases with 4.6% when moving from the N-1 reliability criterion to optimal short-term reliability without optimal investment; and it increases with an additional 12.1% when also investment is optimal. In this case only a quarter of potential surplus gains is obtained when moving to optimal short-term reliability management without altering investment practices. The remaining three quarters are only obtained when the TSO also moves to investment that makes an optimal trade-off of congestion, reliability and investment costs. This division is different for other parameter values. For example, with a low cost of blackout or a low failure probability, the bulk of the surplus gain is obtained by having the optimal short-term trade-off. For the price difference and the supply and demand slopes surplus gains are also substantial but small relative to the change of surplus with these parameters.

In this simple numerical illustration, the gains of moving from N-1 reliability management to optimal short-term reliability management are between 0% and a few tens of percents, and between 14% and a few tens of percents when moving to optimal short- and long-term reliability management.[22] This illustration thus indicates the magnitude of different degrees of optimal reliability decisions. To get a better estimate of the inefficiencies of the N-1 reliability criterion in operational planning and in investment, detailed engineering studies are needed. They model in more detail the network, the failure contingencies, the supply and demand conditions and the values of lost load of different consumer groups at different times. As a first illustration, Heylen et al. (2016b) estimate the short-term efficiency improvement to be 2% for a 5-node system and 6.6% for the 24-bus IEEE-RTS system (Grigg et al., 1999). Second, He et al. (2010) estimate the efficiency improvement to be between 0.3% (normal weather) and 7.1% (adverse weather) for a 6-bus system, and 4.8% for the 24-bus IEEE-RTS. These are short-term efficiency gains. Our numerical illustration indicated

[22]The gains of more optimal decisions are a combination of lower interruption costs, lower investment costs and lower congestion costs. All three gains lead to a higher consumer surplus, either directly through lower wholesale prices or lower interruption costs, or indirectly through lower transmission tariffs.

that optimal investment practices could double or triple these gains.

Figure I.6: Optimal surplus (solid), surplus for optimal short-term reliability (dashed) and surplus for N-1 reliability (dotted) for different values of V_{BO}, p_f, c and $b_E + b_W$.



I.5.2 More complex networks

The previous section illustrated the analysis of this paper in a simple network. It showed that scheduled flow and maximum capacity under the N-1 reliability criterion differ from the optimal values, which are determined from the trade off between congestion, reliability and investment. As this simple network neglected loop flows, this section extends the illustration of optimal reliability and investment to more complex networks with lumpy investments and stochastic operational states.

Assuming the same functional forms and parameter values of section I.5.1, only the expected interruption cost function changes between different network topologies. The interconnection benefit function $S(k)$ is the same as in section I.5.1. However, as expected interruption costs influence both the short-term and long-term first-order condition, both optimal scheduled flow k^* and maximum capacity K^* will change.

Suppose we have the four-node network (a) on the top left-hand side of Figure I.7. Each line has the same reactance, except for the line between node E and node a, which has a reactance that is twice as high. A lossless DC Power flow analysis calculates that 60% of flow from E to I will pass by the southern route, while 40% will use the northern route.[23] Assuming that each line has a thermal limit of 1000 MW, maximum capacity K is 1667 MW (1000/0.6).[24] In that case, the southern lines are at their thermal limit of 1000 MW, while the northern lines are at 667 MW. Scheduled flow determined by the N-1 reliability criterion k_{N-1} is 1000 MW, as after a single line failure only one of the two routes can be used. The optimal scheduled flow is determined from the short-term trade-off between marginal interconnection benefit and marginal expected interruption cost:

$$S'(k^*) = EIC'_k(k^*, K) \tag{1.23}$$

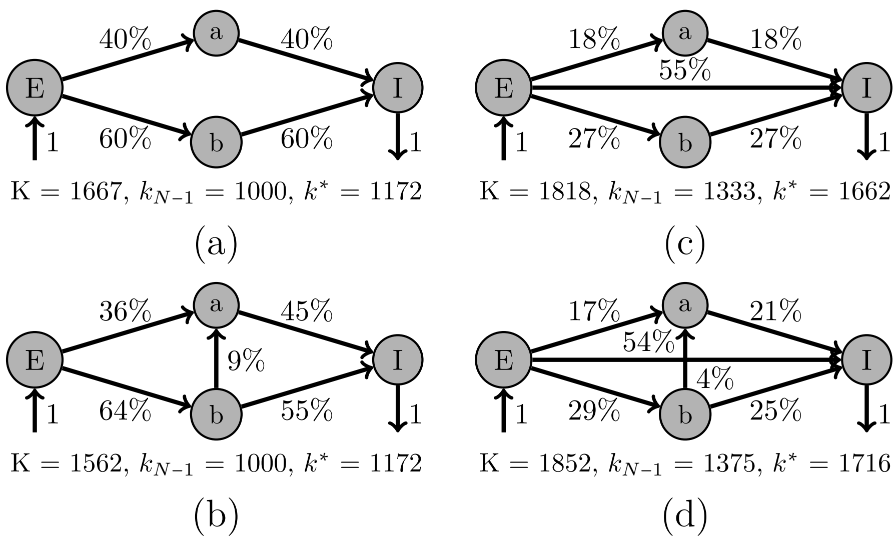
The expected interruption cost $EIC(k, K)$ is calculated using equation (1.21), with the difference that (i) the overflow $(k - \frac{i}{n}K)^+$ has to be evaluated for each individual line, as in the general network formulation of equation (1.14), and (ii) we take the weighted sum over all 2^n contingencies that can happen with n transmission lines: the $\binom{n}{1}$ single line contingencies, the $\binom{n}{2}$ simultaneous double contingencies, etc. Using the parameter values of Table I.1, the optimal scheduled flow k^* is 1172 MW.

Next, assume the same network topology but with an additional line between node a and b. Figure I.7 (b) presents the resulting power flows, which show that the flow through line E-b increases from 60% to 64% of scheduled flow. This phenomenon is known as the Braess Paradox (Braess, 1968) and occurs when a so-called Wheatstone bridge is added between two parallel lines. In that case, more flow passes through the low-reactance line

[23] In short, the higher the reactance x of a path, the less power flows through it. As the reactance of the northern route is $3x$ and the reactance of the southern route is only $2x$, 40% ($\frac{2x}{2x+3x}$) of the flow passes by the northern route.

[24] If the flow on a line l is x_l of total flow and this line's thermal limit is L_l, maximum capacity is determined as the lowest value of $\frac{L_l}{x_l}, \forall l$.

Figure I.7: Capacities K, k_{N-1} and k^* [MW] for different network topologies.



(E-b) to bypass the high-reactance line (E-a) via the Wheatstone bridge. As the maximum line flow in the network increases, maximum capacity decreases to 1562 MW (1000/0.64). Because still only 1000 MW can be transmitted in case line a-I or line b-I fails, the additional Wheatstone line has no reliability benefit in this particular topology and both k_{N-1} and k^* do not change. In less stylized topologies, however, the Wheatstone line has a reliability benefit. For example, when lines a-I and b-I have a larger thermal limit than lines E-a and E-b (e.g. 2000 MW instead of 1000 MW), adding the Wheatstone line increases the N-1 scheduled flow from 1000 MW to 1333 MW (1000/0.75),[25] or, more generally, decreases expected interruption costs at equal scheduled flow:.

$$EIC_{with-ab}(k,K) < EIC_{without-ab}(k,K) \quad \forall k \leq K \tag{1.24}$$

Because of the optimal short-term trade-off (1.23) between congestion and reliability, every improvement of the EIC function automatically transfers to both improved reliability and congestion. When lines E-a and E-b have

[25] The maximum flow of 75% occurs when line a-I fails. In that case, the reactance of the path E-a-b is 3x, while path E-b has a reactance of x.

a thermal limit of 2000 MW instead of 1000 MW, optimal scheduled flow k^* increases from 1172 MW to 1350 MW.

The effect of a Wheatstone addition on reliability and congestion has also been studied by Blumsack et al. (2007). However, by not allowing for a reliability margin (i.e. K=k), they conclude that a Wheatstone line addition increases congestion, because K decreases.[26] By contrast, the above analysis shows that by introducing TSO reliability management into the model, whether it is the N-1 reliability criterion or the optimal reliability criterion, both N-1 scheduled flow k_{N-1} and optimal scheduled flow k^* increase, even though maximum capacity K decreases. The Braess Paradox is thus not present anymore when reliability management is considered.

Figure I.7 (c) shows the power flows of a topology with a direct connection between E and I. As power flows are now more equally spread in the system, maximum capacity increases. Because of the additional line to node I, both the N-1 scheduled flow and the optimal scheduled flow increase. Figure I.7 (d) shows that when a Wheatstone-like line is added to topology (c), maximum capacity, N-1 scheduled flow and optimal scheduled flow increase even more.

Lastly, we study discrete investment decisions under different reliability criteria. As the N-1 and the optimal reliability criterion lead to different levels of congestion and reliability, the investment decision will also be different. Suppose a TSO wants to assess if the two line additions of the right-hand side of Figure I.7 improve net interconnection surplus. Table I.2 shows the change of short-term net surplus, i.e. interconnection benefit minus expected interruption costs, as a function of the price difference Δp. The first part of this table shows results for optimal reliability management. They show that both scheduled flow and short-term net surplus increases with the price difference. The increase is larger for the first investment than for the second.

The second part of Table I.2 shows results for N-1 reliability management. N-1 scheduled flow increases from 1000 MW to 1333 MW for line E-I investment and further to 1375 MW for the additional line b-a invest-

[26] In addition, they show that this trade-off is only present in systems with an embedded Wheatstone sub-network.

Table I.2: Change of scheduled flow [MW] and short-term net interconnection surplus [€/h] for line investment E-I and for an additional line investment b-a, at different levels of Δp. The first part shows optimal results; the second part shows N-1 results.

Δp	k^* [MW]	$S(k^*) - EIC(k^*, K)$ [€/h]
5	1000 → 1360 → 1470	3500 → 4003 → 4062
7	1090 → 1510 → 1590	5586 → 6922 → 7147
9	1170 → 1662 → 1716	7847 → 10097 → 10441
11	1260 → 1780 → 1840	10282 → 13524 → 14014
13	1350 → 1830 → 1840	12890 → 17164 → 17694
Average change of ST optimal net surplus		2321 and 330
Δp	k_{N-1} [MW]	$S(k_{N-1})$ [€/h]
5	1000 → 1333 → 1375	3500 → 4666 → 4813
7	1000 → 1333 → 1375	5500 → 7332 → 7663
9	1000 → 1333 → 1375	7500 → 9998 → 10313
11	1000 → 1333 → 1375	9500 → 12664 → 13063
13	1000 → 1333 → 1375	11500 → 15330 → 15813
Average change of ST 'N-1' net surplus		2498 and 335

ment, irrespective of the price difference. N-1 scheduled flow is always lower than optimal scheduled flow, which varies with the price difference. For N-1 reliability management, only the increase of interconnection benefit matters for the investment decisions, not the reliability benefit, as was shown in equation (1.17). The last column presents the perceived benefit of N-1 investment.

Supposing that all five price differences Δp happen with equal probability throughout the year, the average increase of short-term net surplus is 2321 €/h for the optimal reliability criterion and the perceived short-term net surplus is 2498 €/h for the N-1 reliability criterion. If the cost of the line E-I investment is between 2321 €/h and 2498 €/h, line E-I is installed when the TSO uses N-1 reliability management but not when its reliability management is optimal. Above a cost of 2498 €/h, even a TSO that uses the N-1 reliability criterion will not invest in this line; below a cost of 2321 €/h, all TSOs will invest, irrespective of their reliability management. Line b-a will only be built for very low investment costs.

I.6 Moving beyond the N-1 reliability criterion

The N-1 reliability criterion has been carried over from the old regime of regulated vertically integrated monopolies (Joskow, 2006). It has achieved acceptable results over the past decades, but is considered inadequate in the future system characterized by more decentralized decision makers, more uncertainty, more interconnected networks, more variable renewable generation, more NIMBY[27] and environmental concerns, and a general trend towards a more efficient management. In such a system, the probabilistic approach of this paper is more efficient than the N-1 reliability criterion because it allows TSOs to base their reliability decisions on VOLL, demand, weather conditions, (expected) intermittent generation, etc.

TSOs are starting to be aware of the inefficiencies of the N-1 criterion but many barriers still must be overcome. First, in actual large, meshed networks, calculating the expected interruption cost requires large computing power. Second, calculating the expected interruption cost is a complex

[27]Not In My Back Yard: opposition by local residents to nearby development projects.

issue, while the N-1 criterion is a straightforward and easily comprehensible decision rule. Third, a large amount of technical and economic data is needed. On the technical side one needs demand data, forecast errors, maintenance planning, repair times, wind and solar data, failure probabilities of all system components as a function of temperature and weather[28], etc. On the economic side one needs accurate estimates of interconnection benefit and VOLL.[29]

Despite the above-mentioned barriers, some steps are possible towards a more efficient reliability management. First, instead of only considering single outages, the contingency list could include simultaneous outages with a high probability of occurrence or high-impact outages. Similarly, single outages with very low probability of occurrence or very low impact could be excluded. The decision to include or exclude certain contingencies is made by calculating (an approximation of) the average impact. The contingency list could also depend on the weather and the region. Second, because of advances in communication and information technologies, technical data can be recorded at a decreasing cost. Devices to measure climatic data, real-time voltage and current data, and regional demand and generation data are in an adoption phase in Europe (GARPUR, 2015). Third, VOLL data need to be improved. VOLL studies are laborious and rather expensive, but widespread roll-out of smart meters will facilitate the determination of VOLL of different consumers at different times by offering reliability contracts to (large) consumers. Fourth, TSO and market data could be used to determine costs of preventive and corrective actions – such as the cost of congestion and the cost of balancing and reserves. These can be used to make a trade-off between the costs and benefits of reliability decisions.

The optimal reliability rule always weakly increases net interconnection surplus but this gain has to be weighed against the additional cost of getting

[28]9 out of the 10 most risky days in 2010-2014 in the North American bulk power system were caused by adverse weather (NERC, 2015)

[29]Because of difficulties to estimate VOLL (CEER, 2010), VOLL is usually assumed to be a constant value per TSO zone. In reality the VOLL depends on the type of interrupted consumer, the duration and region of interruption, the time of occurrence, etc. Including these factors yields even higher efficiency gains. Ovaere et al. (2016) estimate additional efficiency gains up to 43%.

the needed data, running a larger simulation, and even the cost of dealing with decreased equity of reliability between consumers. More detailed studies are needed to do this.

I.7 Conclusion

This paper presents a model that explicitly considers the possibility of transmission line outages and uncertainty of demand and intermittent supply. The simple model of this paper clarifies reliability issues in electricity networks. First, there is no direct link between investment and congestion, because of reliability concerns. The effect of line investment on congestion and reliability depends on the reliability criterion. Second, because TSOs keep a reliability margin, the Braess paradox disappears. Line investment increases reliability and decreases investment, for both the N-1 an the optimal reliability criterion. Third, the currently-used N-1 reliability criterion is suboptimal. A probabilistic criterion is more efficient, as it takes into account the expected cost and likelihood of failures. The model shows that the optimal transmission reliability margin is determined by a trade-off between congestion and reliability. These depend on economic and technical parameters and on the network topology, while the currently-used N-1 reliability criterion determines the reliability margin based on the network topology only. Our optimal probabilistic approach is a benchmark that shows the possible efficiency improvements of moving towards a reliability criterion based on expected interruption costs.

This paper provides a qualitative analysis of transmission reliability and reliability criteria. To support the transition towards a probabilistic approach, more quantitative analysis is needed, especially on the effects of the reliability criterion on transmission investment.

In a system characterized by increased renewable generation and difficulties to build new lines, power system operation will be closer to its transmission capacity limits. In such a stressed system the gains of a probabilistic reliability management are expected to be even higher. Fortunately, advances in communication and information technologies are making it possible to move towards a probabilistic approach of reliability management.

A reliability criterion that is adapted to the challenges and needs of our modern society affects all of us. It allows a better assessment of the risks and better integration of renewable generation, while lowering the cost of our electricity system.

An important caveat is that TSOs are assumed to pursue the correct efficiency objective and that TSOs cooperate in international transmission to maximize their joint surplus. This is obviously more difficult to monitor when TSOs use a probabilistic reliability criterion than when they use the N-1 reliability criterion.

Acknowledgements

We are grateful for very helpful comments from three anonymous reviewers and Fridrik Mar Baldursson, Claude Crampes, Thomas-Olivier Léautier, Machiel Mulder, Guido Pepermans, Dirk Van Hertem, Frank Verboven and Bert Willems. Work is conducted through funding by the GARPUR project of the 7th European Community Framework Programme: http://www.garpur-project.eu/.

I.A Power flow calculation

The solution of a power flow is usually found using numerical methods (Zimmerman et al., 2011), but it is possible to manually calculate power flow for simple networks using the superposition method. As an illustration, this appendix explains the power flow calculation of the four-node network of section I.4.1 where line $W_1 - E_1$ is out. Figure I.8 shows the four different ways electricity flow is transferred from West to East when the direct line between W_1 and E_1 is out: $W_1 \rightarrow E_1$, $W_1 \rightarrow E_2$, $W_2 \rightarrow E_1$ and $W_2 \rightarrow E_2$. For each of the cases one can calculate how the flow is distributed over the different lines. The flow distribution is determined by the reactance of the difference paths. In short, the higher the reactance of a path, the less power flows through it. As before, suppose that all lines have an identical reactance x.

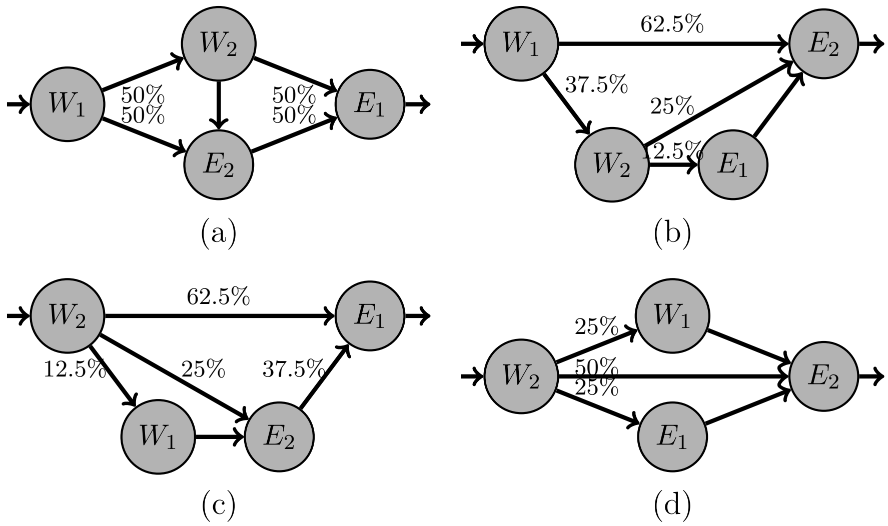


Figure I.8: Manual power flow calculations.

In case (a) the path through W_2 and the path through E_2 have equal reactance $2x$, such that the flow is distributed equally over both paths. In case (b) the path $W_1 - W_2 - E_1 - E_2$ has a reactance of $1 + \frac{1}{1+\frac{1}{1+1}} = \frac{5}{3}$, while the direct path between W_1 and E_2 has a reactance of $x = 1$. Therefore, the flow on line $W_1 - E_2$ is $\frac{5/3}{1+5/3} = 62.5$ %, while the flow via path $W_1 - W_2 - E_1 - E_2$ is 37.5%. The flow distribution of cases (c) and (d) are calculated in the same way.

The final flow on a line is calculated as the superposition of the four cases. For example, the flow on line $W_1 - W_2$ is $0.25(0.5+0.375-0.125-0.25) = 12.5$ %, while the flow on line $W_1 - E_2$ is $0.25(0.5+0.625+0.125+0.25) = 37.5$ %.

Chapter II

A Multi-Dimensional Analysis of Reliability Criteria: From Deterministic N-1 to a Probabilistic Approach

II.1 Introduction

A large number of papers have been studying probabilistic reliability criteria as alternatives to the currently-used deterministic N-1 reliability criterion (Karangelos and Wehenkel, 2016a; He et al., 2010; Fu and McCalley, 2001; Capitanescu et al., 2007, 2011a, 2012; Dogan et al., 2016). They argue that probabilistic reliability criteria are better suited to meet the current challenges of the electricity transmission system: uncertain and variable demand and supply, decentralized decision makers, highly interconnected networks, difficulties in building new lines and a general trend towards a more efficient use of the transmission system (Ovaere and Proost, 2016). Because probabilistic reliability criteria make a trade-off between costs and benefits of reliability actions, they lower system costs, especially in stressed and volatile systems.

However, despite their advantages, probabilistic reliability criteria are not yet used by transmission system operators (TSOs)[1] in operational planning and real-time operation. TSOs are not eager to change their reliability management and criterion, due to both the satisfactory results obtained in the past and the transparency of the currently used N-1 reliability criterion. Probabilistic reliability criteria, in contrast, are perceived to be more complex and less transparent.

This paper argues that the strict dichotomy between the deterministic N-1 criterion and a full probabilistic approach is a simplification. Several reliability criteria exist between these two extremes. To distinguishes intermediate steps, this paper proposes a classification of reliability criteria based on four characteristics: (i) the set of considered system states, (ii) the objective function, (iii) the allowed real-time actions and (iv) optional non-technical constraints. Next, in order to convince TSOs to move away from the deterministic N-1 criterion, this paper suggests an assessment of different reliability criteria along five dimensions: (i) expected total cost, (ii) unreliability, (iii) inequality between consumers, (iv) data needs and availability and (v) ease of use. Finally, we illustrate the multi-dimensional performance assessment of our six proposed reliability criteria in a five-node test system. This illustration, however, does not intend to identify the fundamentally optimal reliability criterion for actual large-scale systems, but intends to indicate general characteristics and performance of different reliability criteria. TSOs and regulators can use the methods of this paper to carry out a multi-dimensional analysis for their own system.

Section II.2 presents the classification and describes six reliability criteria that range from the deterministic N-1 reliability criterion to a full probabilistic reliability criterion. Section II.3 discusses the five indicators to evaluate the performance of reliability criteria. The six reliability criteria are then evaluated along the five indicators in a case study in section II.4.

[1]We use the term 'TSO' throughout this paper, but the core of our analysis does not change with the level of ownership unbundling of the network (full ownership unbundling, independent system operator or independent transmission operator) or with the geographical area of focus (e.g. in North America: independent system operator or regional transmission organization)

Table II.5 summarizes the result of this multi-dimensional analysis. Section II.5 discusses the results, while section II.6 discusses a practical application of the classification framework. Section II.7 concludes.

II.2 Reliability criteria

Reliability criteria guide the reliability management of TSOs, from long-term system development to short-term operational planning and real time operation (Ovaere and Proost, 2016). In each of these planning horizons, the TSO continuously takes actions to minimize the cost of satisfying the reliability criterion. This paper focuses on the minimization problem of the TSO in operational planning and real-time operation:

$$\min_{a_p, a_c, P_c} [C(a_p, a_c, P_c)] \tag{2.1}$$

$$\text{s.t. } G_s(a_p, a_c, P_c) = 0, H_s(a_p, a_c, P_c) \geq 0 \quad \forall s \in S \tag{2.2}$$

In operational planning and real-time operation, the TSO's objective is to minimize total cost $C(a_p, a_c, P_c)$, while satisfying the power flow equations (equality constraints G_s) and operational limits (inequality constraints H_s) (Capitanescu et al., 2011b). During operational planning, the TSO takes the most cost-effective preventive actions a_p to ensure that these constraints are met in all considered system states $s \in S$.[2] The set of considered system states also depends on the reliability criterion.

If contingencies happen in real time and preventive actions turn out to be insufficient, the TSO can take corrective actions a_c or resort to load curtailment P_c.[3] They choose the cheapest actions that are within the constraints of the applied reliability criterion to make sure operational limits are still met. Unconsidered system states could lead to uncontrolled brownouts or blackouts, if corrective actions are not able to deal with the realized real-time system state.

[2]Examples of available actions in the operational planning stage are generation re-dispatch, branch switching, phase shifting transformer tap changing, and ensuring the availability of upward and downward reserves (Baldursson et al., 2016b).

[3]Possible corrective actions are branch switching, secondary voltage control, capacitor and reactor bank switching, the use of upward and downward reserves, phase shifting transformer tap changing and load curtailment (Capitanescu et al., 2011b).

The remainder of this section formulates and discusses six reliability criteria.[4] They are defined by four characteristics: (i) the set of considered system states S, (ii) the allowed TSO actions in considered system states, (iii) the objective function to minimize and (iv) optional non-technical constraints. Table II.1 summarizes the analyzed reliability criteria along these four characteristics.[5]

II.2.1 N-1 reliability criterion

Currently, all TSOs use the N-1 reliability criterion or some variant in short term reliability management. This straightforward criterion states that after an unexpected outage of a single system component, the transmission system should still be able to accommodate all flows without load curtailment. A direct link between the preventive and corrective stage is not made if not required and the system is secured ahead of real time, if possible. The TSO's objective function is deterministic and limited to minimizing the cost of preventive actions. The expected cost of corrective and curtailment actions in real time are not explicitly considered. The set of considered system states consists of all N-1 contingencies and is usually called the N-1 contingency set. In case of N-1 network contingencies, the network should be able to accommodate all resulting flows. The mathematical formulation of the N-1 reliability criterion is:

$$\min_{a_p} \; [C_{prev}(a_p)] \tag{2.3}$$

$$\text{s.t. } G_{s_n}(a_p) = 0 \quad \forall s_n \in S_{N-1,network} \tag{2.4}$$

In case of N-1 generation contingencies, real time corrective actions a_c, like upward and downward use of reserves, are needed to restore the balance between demand and supply.

$$G_s(a_p, a_c^s) = 0 \quad \forall s \in S_{N-1,generation} \tag{2.5}$$

[4] To simplify the notation, $G_s()$ represents both the equality and inequality constraints in system state s.

[5] The discussed reliability criteria mainly focus on risk-neutral reliability management. Alternatively, risk averse objective functions can be considered that are typically implemented using robust optimization techniques (Bertsimas et al., 2011).

In any case, load curtailment is not allowed in N-1 system states (i.e. in both network and generation N-1 contingencies). The N-1 reliability criterion does not explicitly prepare for multiple contingencies. In these cases, load curtailment could turn out to be required to prevent a blackout in real time operation.

In many countries, the exact definition of the N-1 reliability criterion differs from the above strict formulation. The next reliability criterion allows for a more general set of considered system states than the strict N-1 contingency set.

II.2.2 Deterministic reliability criterion with a different set of considered system states

The mathematical formulation of this reliability criterion is similar to that of the N-1 reliability criterion. the primary difference being that the set of considered system states is allowed to differ from the N-1 contingency set.

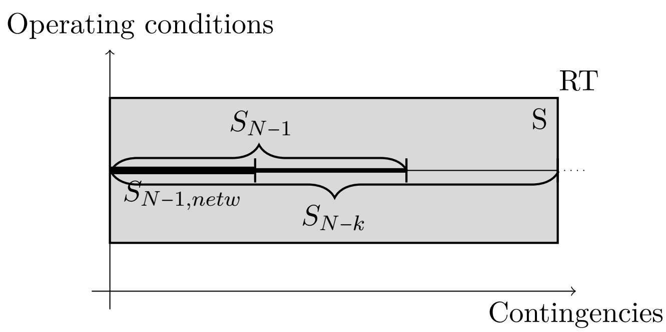
$$\min_{a_p} \; [C_{prev}(a_p)] \tag{2.6}$$

$$\text{s.t. } G_s(a_p) = 0 \quad \forall s \in S_p \tag{2.7}$$

$$\text{s.t. } G_s(a_p, a_c^s) = 0 \quad \forall s \in S \setminus S_p \tag{2.8}$$

This reliability criterion requires a TSO to minimize its cost of preventive actions C_{prev} while meeting the constraints for a subset of all considered system states S_p with preventive actions only and for the remaining considered system states $S \setminus S_p$ with both preventive and corrective actions in order to balance demand and supply. The set of considered system states S could be anything. For example, include N-1 network contingencies, but exclude generator and busbar failures from the N-1 contingency set (GARPUR, 2014, p.25); increase the contingency set to include multiple dependent failures with a high probability of occurrence; change the contingency set over time (e.g. including double-circuit failures only during adverse weather) or between regions (e.g. including more contingencies for urban areas or business districts); etc. In addition to contingencies, the set of considered system states could also include deviations from the expected operating condition, like forecast errors of demand and intermittent supply. In its most general form, the set of considered system states S is the Cartesian product of credible contingencies and considered operating conditions. S is always a subset of the infinite set of all possible future real-time states and contingencies (RT). Figure II.1 illustrates different sets of considered system states in the Cartesian plane of contingencies and operating conditions: the set of N-1 network contingencies $S_{N-1,netw}$, the set of N-1 contingencies S_{N-1}, a set of N-k contingencies S_{N-k}, and a general set S of contingencies and operating conditions.

Figure II.1: Different sets of considered system states in the Cartesian plane of contingencies and operating conditions.



It is also possible that (well-informed) state selection leads to a set of considered system states in which not all of the N-1 contingencies are included. This might be, for instance, because the probability of occurrence or the impact of the excluded states is too low. The impact on the performance indicators of such sets is hard to predict, as this is an interaction between the impact of the additional and removed operating states.

II.2.3 Probabilistic reliability criterion without load curtailment in considered states S

If the TSO takes the expected cost of corrective actions in considered system states $s \in S$ into account in operational planning, its objective function becomes probabilistic. The TSO simulates which actions it will take in each of the considered system states and the expected cost is calculated as the product of the cost of actions in each state and its associated probability of occurrence. First, suppose that load curtailment is not allowed in considered system states. The mathematical formulation becomes:

$$\min_{a_p, a_c^s} \left[C_{prev}(a_p) + \sum_{s \in S} \pi_s \left(C_{cor}(a_c^s) \right) \right] \tag{2.9}$$

$$\text{s.t. } G_s(a_p, a_c^s) = 0 \quad \forall s \in S \tag{2.10}$$

The difference between this criterion and the previous criterion is that a TSO now incorporates the effect of its preventive actions on the cost of its corrective actions, instead of just checking if the constraints are met. This allows an explicit trade off between preventive and corrective actions.

II.2.4 Probabilistic reliability criterion

If, in addition, load curtailment is allowed in considered system states and the TSO also takes the expected cost of load curtailment into account in its operational planning minimization, this results in a full probabilistic reliability criterion (McCalley et al., 2004; Kirschen and Jayaweera, 2007; He et al., 2010). The mathematical formulation is:

$$\min_{a_p, a_c^s, P_c^s} \left[C_{prev}(a_p) + \sum_{s \in S} \pi_s \left(C_{cor}(a_c^s) + C_{curt}(P_c^s) \right) \right] \tag{2.11}$$

$$\text{s.t. } G_s(a_p, a_c^s, P_c^s) = 0 \quad \forall s \in S \tag{2.12}$$

The difference between this criterion and the previous criterion is that the effect of preventive actions on the cost of corrective actions and load curtailment is incorporated. This allows an explicit trade off between preventive, corrective and curtailment actions.

II.2.5 Probabilistic reliability criterion with a constraint on the aggregate reliability level

While the full probabilistic reliability criterion aims at minimizing the expected total cost (ETC), it can also reduce the reliability level considerably (Ovaere et al., 2016). Therefore, social and political concerns could lead to the addition of a constraint on the value of the aggregate reliability level $\bar{P}$.

Such a constraint limits the decrease of the reliability level. If the constraint is binding, ETC will be higher. The mathematical formulation becomes:

$$\min_{a_p, a_c^s, P_c^s} \left[C_{prev}(a_p) + \sum_{s \in S} \pi_s \left(C_{cor}(a_c^s) + C_{curt}(P_c^s) \right) \right] \tag{2.13}$$

$$\text{s.t. } G_s(a_p, a_c^s, P_c^s) = 0 \quad \forall s \in S \tag{2.14}$$

$$\text{s.t. } \sum_{j \in J} \sum_{s \in S} \pi_s P_c^s(j) \leq \bar{P} \tag{2.15}$$

with J the set of all consumers and $P_c^s(j)$ the load curtailment [MW] of consumer j in state s.

II.2.6 Probabilistic reliability criterion with constraints on individual reliability levels

The reliability constraint can also be imposed at the level of the individual instead of the aggregate. In that case, the constraint provides a minimal reliability level $\bar{P}_c(j)$ for each consumer j.

$$\min_{a_p, a_c^s, P_c^s} \left[C_{prev}(a_p) + \sum_{s \in S} \pi_s \left(C_{cor}(a_c^s) + C_{curt}(P_c^s) \right) \right] \tag{2.16}$$

$$\text{s.t. } G_s(a_p, a_c^s, P_c^s) = 0 \quad \forall s \in S \tag{2.17}$$

$$\text{s.t. } \sum_{s \in S} \pi_s P_c^s(j) \leq \bar{P}_c(j) \quad \forall j \in J \tag{2.18}$$

II.3 Performance evaluation of reliability criteria

The performance of power system reliability criteria is multi-faceted and several opposing objectives need to be considered. Some of them can be quantified, while others are qualitative in nature.

II.3.1 Quantitative indicators

There are three important quantitative indicators that determine the performance of different reliability criteria: expected total cost (ETC), relative load curtailment (RLC) and inequality between consumers (U_{ENS}).

Table II.1: Summary of reliability criteria.

	(a)	(b)	(c)	(d)	(e)	(f)
1. Set of considered states	S_{N-1}	S_{N-k}	S	S	S	S
2. Curtailment allowed in S	no	no	no	yes	yes	yes
3. Objective function	Det.[1]	Det.	Prob.[2]	Prob.	Prob.	Prob.
4. Non-technical constraints	/	/	/	/	$\bar{P}$	$\bar{P}_c(j)$

[1] Deterministic [2] Probabilistic

(a) Deterministic with N-1 contingency set
(b) Deterministic with different set of considered states
(c) Probabilistic without curtailment
(d) Probabilistic
(e) Probabilistic with aggregated constraint
(f) Probabilistic with individual constraint

Socio-economic performance is evaluated in terms of expected total cost. It is the sum of the cost of preventive actions, the cost of corrective actions and interruption costs. Interruption costs equal the amount of load curtailed times the value of lost load (VOLL) and represent the consequences of an interruption for the consumers. The total cost (TC) at a certain time t and real-time state rt is equal to:

$$\begin{aligned} TC(rt,t) = \sum_{a_p \in A_p} C_{prev}(a_p) + \sum_{a_c \in A_c} C_{cor}(a_c) \\ + \sum_{j \in J} VOLL(j) \cdot P_c(j) \end{aligned} \tag{2.19}$$

where A_p is the set of executed preventive actions, A_c the set of executed corrective actions and J the set of consumers in the system. The expected total cost (ETC) over all times t and real-time states rt is equal to:

$$ETC = \sum_{t \in T} \pi(t) \sum_{rt \in RT} \pi(rt|t) \cdot TC(rt,t) \tag{2.20}$$

with $\pi(t)$ the probability of being at time t and $\pi(rt|t)$ the probability of being in real time state rt at time t. The set T contains for example all 8760 hours of a typical year or a representative subset of time instants. Since the set RT is the infinite set of all possible contingencies and all possible operating conditions, (2.20) is in practice evaluated for a finite subset $RT' \subseteq RT$, where the set of considered system states $S \subseteq RT'$ (Heylen et al., 2016a).

The amount of load curtailment in the system, aggregated, per node or per consumer, is here expressed in terms of relative load curtailment. Relative load curtailment RLC is rescaled to the equivalent number of minutes curtailed in a year:

$$RLC = \left(1 - \frac{P_D - P_c}{P_D}\right) \cdot 8760 \cdot 60 \qquad [\text{min/year}] \tag{2.21}$$

with P_D the total demand and P_c the curtailed load. The indicator RLC is thus a measure of the average unreliability of the system.

Reliability criteria that result in large differences of RLC between consumers or between nodes might not be socially acceptable. Therefore, equality is another important aspect to consider in the selection of a practical reliability criterion. Inequality can be quantified in a single number by using a Gini-based performance indicator U_{ENS}. The inequality coefficient U_{ENS} is defined as (Heylen et al., 2017a):

$$U_{ENS} = |1 - (\sum_{k}(D_k - D_{k-1}) \cdot (E_k + E_{k-1})| \tag{2.22}$$

where D is the cumulative share of demand, E the cumulative share of energy not supplied (for example over a year) and k an index counting over consumers or consumer groups. The groups are ordered based on decreasing reliability values. If U_{ENS} equals 0, unreliability is distributed equally among the entities under study, i.e. nodes, consumer groups or individual consumers.[6] If the inequality indicator U_{ENS} is closer to 1, all interruptions are concentrated in one or a few consumer groups or nodes (Ovaere et al., 2016).

II.3.2 Qualitative indicators

Next to the three quantitative performance indicators, two qualitative indicators are considered: (i) the amount and availability of the required input data and (ii) ease of use. The first refers to the complexity, correctness and cost of acquiring input data; while the second refers to the ease of interpreting and reacting to the output. As these are hard to quantify, a scoring system from +++ to - is used.

Appropriate data are important for decent reliability management. The amount and availability of the required input data are qualitative aspects that should be considered when evaluating alternative reliability management strategies. Handling a large amount of possibly more complex data makes the methodology more complex and the results can be sensitive to the correctness of the provided data. Moreover, collecting additional data is time consuming and can be costly. An additional qualitative indicator that represents the effect of these aspects should be considered in the criterion selection.

The ease of use of the currently used N-1 criterion is one of the reasons why TSOs are not eager to change their reliability management. Proba-

[6] Note that the inequality indicator can not be calculated if reliability is 100% for all consumer groups in all nodes. In that case inequality U_{ENS} equals 0.

bilistic approaches are typically risk based, taking into account exact probabilities and severities of the considered operating states. Deterministic approaches, on the other hand, simplify the situation by assuming all considered operating states to be equally probable and equally severe. Moreover, ease of use is influenced by the number of operating states to consider. Given the increased uncertainty in power systems due to the increased penetration of renewable energy sources, it might be beneficial to consider additional operating states in the decision making. However, this makes the decision making process more complex.

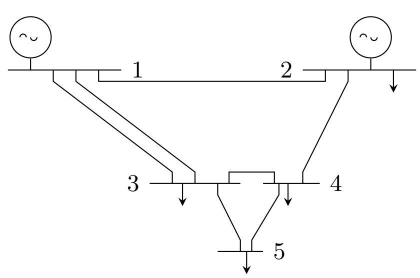
II.4 Case study

A case study on a five-node test system applies the proposed performance evaluation to the six reliability criteria introduced in section II.2. The results are summarized in Table II.5.

II.4.1 Data

The illustrative five-node test system is based on the Roy Billinton reliability test system (Billinton et al., 1989), as shown in Fig. II.2. Transmission line and generation data are summarized in tables II.2 and II.3 respectively.

Figure II.2: Circuit diagram of the test system



Total system demand is based on the hourly load profile defined in (Grigg et al., 1999). For simplification a year is represented by 72 time instants, each with its probability of occurrence.[7] Total system demand at

[7]The 72 time instants represent 6 periods in the year (winter, early spring, late spring, summer, early autumn and late autumn), 3 types of days (weekday, Saturday and Sunday)

Table II.2: Line data

From node	To node	x [pu]	Capacity [MVA]	Failure probab.
1	3	0.18	85	0.0017
2	4	0.6	71	0.0057
1	4	0.48	71	0.0046
3	4	0.12	71	0.0011
3	5	0.12	71	0.0011
1	3	0.18	85	0.0017
4	5	0.12	71	0.0011

Table II.3: Generation data

Node	Capacity [MW]	Type	C_{marg} [€/MWh]	Failure probab.
1	40	coal	13.83	6.2E-7
1	40	coal	13.83	6.2E-7
1	10	coal	13.83	6.2E-3
1	20	wind	0.04	6.2E-3
2	40	coal	13.83	6.2E-7
2	20	coal	13.83	6.2E-3
2	20	wind	0.01	6.2E-3
2	20	wind	0.03	6.2E-3
2	20	wind	0.05	6.2E-3
2	5	coal	13.83	6.2E-3
2	5	coal	13.83	6.2E-3

Table II.4: Summary of the three cases for the sensitivity analysis

	More stressed	Base case	Less stressed
Load	105%	100%	95%
Failure rates	150%	100%	75%
Repair times	150%	100%	75%
Line rating	91%	100%	136%

each of the 72 time instants is calculated as the mean over all valid hours. By weighing the outcomes of the different time instants by their probability of occurrence, the quantitative indicators are evaluated for a year.

This numerical illustration uses VOLL data from Norway (EnergiNorge, 2012). Two consumer types are considered: residential and non-residential customers. The share of residential and non-residential demand in total system demand changes throughout the year, as presented in (Ovaere et al., 2016).

In this paper, an analytical non-sequential state enumeration technique is applied. The quantitative performance indicators are evaluated for a set of time instants T for which forecast values for load and renewable power generation are given. Preventive reliability management is simulated for a set RT of real time realizations. This set is the Cartesian product of possible contingencies and real time operating states.[8] The latter are conditional upon the forecast values of net demand. Preventive and corrective reliability management are modeled using a DC security constrained optimal power flow (DC-SCOPF) in AMPL (Fourer et al., 1987), taking into account the specifics per criterion as discussed in the next section. The simulations are done in Matlab using an AMPL interface.

The simulation is repeated for a more stressed and a less stressed case as defined in table II.4 in order to verify the sensitivity of the results.

and 4 times during the day (night, morning, noon and evening).

[8] Most probable contingencies up to cumulative probability of 99.73% and 11 realizations of net total demand based on a normal distribution of which the mean equals the forecast value and the coefficient of variation is 4%.

II.4.2 Implementation of the reliability criteria

The reliability criteria introduced in table II.1 are practically implemented as follows.

Criterion (a) considers all N-1 branch and generator outages in the preventive decision stage. All these operating states are considered to be equally probable. Load curtailment is avoided for this contingency set and all consumers are treated equally. The objective is to secure the system preventively as much as possible and corrective actions are considered as a last resort. The above also holds for criterion (b), but the set of considered contingencies is different.

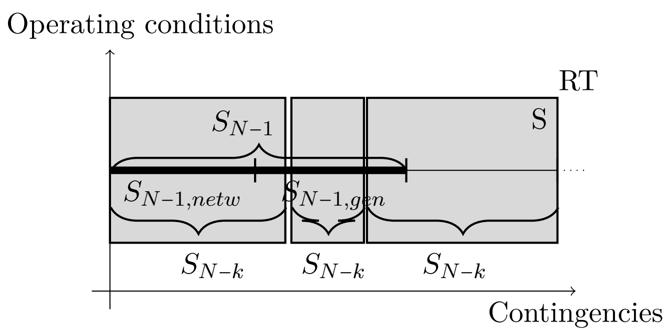
Criteria (b) - (f) consider a larger set of contingencies than criterion (a): The most probable contingencies up to a cumulative probability of 99.7% are considered. Therefore, the failure of some large generator units is not considered in the contingency set of criteria (b) - (f) in this case study, due to their low probability of failure.

The set of considered system states S for reliability criteria (c) - (f) consists of the Cartesian product of the elements of the contingency set and 7 possible real time realizations of net total demand. These realizations are determined based on a normal distribution with the forecast value of net total demand as mean and a coefficient of variation of 4%. Probabilistic criteria (c) - (f) take into account the exact probabilities in the decision making.

While probabilistic criterion (c) tries to avoid load curtailment, criteria (d) - (f) allow load curtailment if this tends to be more cost effective. However, criteria (e) - (f) have upper limits on the amount of load curtailment, i.e. 0.4 MW aggregated load curtailment for criterion (e) and 0.4 MW load curtailment distributed over the consumer groups according to their demand share for criterion (f).

The set of operating states used in criterion (b), S_{N-k}, and criteria (c) - (f), S, is graphically illustrated in Fig. II.3. Some system states included in the N-1 contingency set are not included in the alternative sets S_{N-k} and S, due to their low probability of occurrence

Figure II.3: Graphical representation of the set of operating states used in criterion (1) S_{N-1}, criterion (2) S_{N-k} and criteria (3) - (6) S.



II.4.3 Results

Table II.5 summarizes the performance of the six reliability criteria in the considered case study.

The first row of Table II.5 shows that expected total cost decreases when moving from criterion (a) to criterion (d) and increases again when imposing restrictions on the aggregate (crit. (e)) and individual (crit. (f)) reliability level. Criterion (b) results in a lower ETC than criterion (a), because in our case study the set of considered system states S excludes low-probability contingencies that require costly preventive actions. Generally, the change of ETC from (a) to (b) depends on the current performance of the N-1 criterion. If the N-1 criterion is too stringent, enlarging the set of considered system states leads to even higher total costs. If the N-1 criterion is too loose, enlarging the set of considered system states could lead to lower total costs. Next, ETC decreases by moving from deterministic criteria (a) and (b) to probabilistic criterion (c), because more information is included in the operational planning decision. Criterion (c) makes an explicit trade-off between preventive and expected corrective actions (Strbac et al., 1998). By additionally allowing load curtailment, if this is less costly than alternative corrective actions, criterion (d) leads to even lower ETC. In this case study, most of the decreased ETC is due to a better trade-off between preventive and expected corrective actions (crit. (c)). A smaller part is due to allowing load curtailment in considered system states (crit. (d)). The higher VOLL is compared to the costs of other corrective actions, the larger this effect (Ovaere et al., 2016). The ETC of criteria (e) and (f) can be anywhere between the ETC of criteria (a) and (d), and even higher than the ETC of criterion (a), depending on the level of the constraint. The more stringent the imposed reliability constraint, the higher the ETC. Individual constraints on the reliability level (crit. (f)) always lead to a higher ETC than an aggregate constraint (crit. (e)), as load curtailment of low-VOLL consumers is now substituted for corrective or preventive actions, or load curtailment of consumers with a higher VOLL. To summarize:

$$\begin{aligned} &ETC_d \leq ETC_c \leq ETC_b \text{ and } ETC_a \\ &ETC_f \geq ETC_e \geq ETC_d \end{aligned} \tag{2.23}$$

The second row shows the unreliability for each of the six criteria. The effect of a particular reliability criterion on the reliability level is closely related to its effect on ETC. Criterion (b) has a slightly higher unreliability, because in our case study not all elements in the N-1 contingency set S_{N-1} are part of the set S. Moving from deterministic to probabilistic criteria has a mixed effect on reliability, depending on the exact formulation of the criterion. Disallowing load curtailment in considered states (crit. (c)) leads to an equal or lower unreliability than for deterministic reliability criteria. Unreliability is not zero because load curtailment could still be necessary in non-considered states. Allowing load curtailment in considered states (crit. (d)) decreases ETC even more, but at the expense of a highly increased unreliability. Evidently, constraints on the reliability level (crit. (e) - (f)) decrease the unreliability, but at the expense of a higher ETC. To summarize:

$$RLC_c \leq RLC_d \geq RLC_e \geq RLC_f \tag{2.24}$$

The third row shows the inequality between consumers. It seems that inequality is inversely proportional to ETC. Deterministic criteria lead to higher ETC and lower inequality.[9] Probabilistic criteria lead to lower ETC and higher inequality. The reason is that deterministic criteria treat all consumers equally, while probabilistic criteria differentiate between consumers

[9] If the amount of load curtailment is too low, i.e. RLC < 0.01 min in this case study, the inequality does not give useful information.

or nodes in terms of VOLL. Constraints on the reliability level decrease inequality, but at the expense of higher ETC. Inequality is lower if individual constraints on load curtailment are applied (crit. f) than for criteria (d) and (e). This is the case, because the inequality resulting from the differentiation in VOLL is reduced by these individual limits, which are based on consumers' demand share. To summarize, we observe the following inequalities in our case study[9]:

$$\begin{aligned} &U_{ENS,d} \geq U_{ENS,c} \geq U_{ENS,b} \geq U_{ENS,a} \\ &U_{ENS,f} \leq U_{ENS,e} and U_{ENS,d} \end{aligned} \tag{2.25}$$

The fourth row shows that data needs increase from left to right. Probabilistic approaches (c) - (f) typically require more information than deterministic approaches. Firstly, exact probabilities of failures should be known. Moreover, forecast errors can be considered and for criterion (e) and (f) appropriate limits on load curtailment should be determined. Some of these data are currently not available for TSOs and might be hard to obtain in practice. Failure probabilities might be imprecise, as well as forecast errors, which might lead to inappropriate reliability management.

The last rows show that the ease of using the considered criteria decreases from left to right, because more operating states need to be considered, more information needs to be taken into account and the probabilistic nature of criteria (c) - (f) adds a layer of complexity to reliability management.

Table II.6 summarizes the results of the sensitivity analysis for the two cases defined in table II.4. In the more stressed case, the relative difference in total system cost between the deterministic and probabilistic approaches is slightly decreased. Both in the more stressed and the less stressed case, the trends in differences between the criteria in terms of ETC_{rel}, RLC and U_{ENS} are the same as in the base case.

II.5 Discussion

Once the list of credible operating states against which the system needs to be secured is specified for a deterministic criterion (e.g. crit. (1) and

Table II.5: Performance evaluation of the six considered reliability criteria: The first three indicators give a numerical value and the last two indicators are expressed as a value (-/+/++/+++) from worst to best (relative).

	Criteria of table II.1					
	(a)	(b)	(c)	(d)	(e)	(f)
1. ETC_{rel} [%]	100	90.09	34.40	26.63	33.83	34.50
2. RLC [min]	0.0046	0.0074	0.0046	18.87	1.83	0.19
3. U_{ENS} [/]	0.741	0.3212	0.819	0.811	0.794	0.604
4. Data needs and availability	+++	++	+	+	-	-
5. Ease of use	+++	++	+	+	-	-
5a. Type	Det.	Det.	Prob.	Prob.	Prob.	Prob.
5b. # of states[1]	19	28	196	196	196	196
5c. # info	+++	++	+	+	-	-

[1] Dependent on the state selection algorithm
The numerical values cannot be generalized to real systems, but the trends between reliability criteria can.

Table II.6: Sensitivity of the six considered reliability criteria in a more stressed case and a less stressed case

More stressed case	Criteria of table II.1					
	(a)	(b)	(c)	(d)	(e)	(f)
ETC_{rel} [%]	100	86.38	36.05	27.89	37.04	37.42
RLC [min]	0.0796	0.0689	0.0632	20.3	1.92	1
U_{ENS} [/]	0.353	0.411	0.552	0.815	0.693	0.581
Less stressed case	**Criteria of table II.1**					
	(a)	(b)	(c)	(d)	(e)	(f)
ETC_{rel} [%]	100	90.46	32.31	25.00	31.66	32.33
RLC [min]	0.0002	0.0016	0.0001	17.7	1.82	0.189
U_{ENS} [/]	0.718	0.476	0.620	0.814	0.829	0.629

(2)), it is straightforward to use. Probabilistic reliability management is less transparent and more difficult to handle due to the higher number of system states. Appropriate state selection techniques are required (Platbrood et al., 2011) as the set of considered states impacts the performance. A well selected set of considered system states reduces conservatism regarding low probability N-1 states which may be costly to secure. Additional states with high impact may be considered. The optimal set of operating states for a particular criterion differs between systems.

More advanced probabilistic reliability management approaches require a larger amount of typically more complex data. This makes the methodology more complex. Even a simple probabilistic reliability management strategy requires additional information. Firstly, failure and repair rates of components should be known, which might be imprecise due to the low number of failures of system components. Results can be sensitive to the exactness of the values of the provided data. In this respect, accounting for imprecise probability in the decision making might be favourable (Coolen et al., 2011). Secondly, additional costs are taken into account that are difficult to determine. One of the additional costs that is considered in a probabilistic approach is the interruption cost, which depends on the value of lost load (VOLL). By using detailed VOLL data, we can differentiate between consumers. This leads to large efficiency savings, but comes at the cost of increased inequality of reliability, as shown in table II.5 and (Ovaere and Proost, 2016; Heylen et al., 2017a). Thirdly probability distributions of forecast errors of load and wind power generation can be taken into account, which might be multivariate in nature.

In practice, additional decision stages exist in short term reliability management on top of the two decision stages considered in this paper ()i.e. day ahead operational planning and real time operation). One distinguishes the D-2 decision stage, the intraday decision stage with the intraday market, the short term preventive stage and the short term stage in which automatic actions are performed. Corrective actions which are required in real time in order to secure the system might not work as expected. This is an additional uncertainty that can be considered in probabilistic reliability management (Karangelos and Wehenkel, 2016a; Vadlamudi et al., 2016;

Calvo et al., 2016). Each of these decision stages is also influenced by external factors that are out of the control of the TSO, such as markets, balance responsible parties, generators and loads.

As the five performance indicators have different units, an optimal reliability criterion does not exist. It is the regulator's task to determine the adequate reliability criterion based on the TSO's capabilities and society's preferences.

II.6 Application of the classification framework

Probabilistic reliability criteria are not yet used in practice, but the EU FP7 project GARPUR has proposed and analyzed an advanced, probabilistic, short-term reliability management approach and criterion (Karangelos and Wehenkel, 2016a; GARPUR, 2016). The GARPUR approach can be analysed by our proposed classification framework along the four characteristics of Table II.1:

1. The set of considered system states is determined using a discarding principle, which neglects a subset of contingencies for which the expected interruption cost is lower than a specified maximal, residual risk level. This set differs for different time instants.
2. Curtailment is allowed in considered system states.
3. The objective function is probabilistic and aims at minimizing total socio-economic system cost.
4. A reliability target ensures that the probability of reaching non-considered system states is lower than a fixed tolerance. This non-technical constraint effectively puts a limit on the aggregate reliability level.

In light of these four characteristics, the reliability approach proposed in the GARPUR project resembles criterion (e) of our classification framework.[10] As analyzed in section II.4.3, this category of criteria leads to a lower ETC than deterministic criteria, without overly decreasing reliability.

[10] While this paper focuses on the four high-level characteristics of different reliability criteria, the GARPUR project has focused more on the equally-important topic of the exact implementation of probabilistic criteria and the tuning of decision parameters, such as the set of considered system states (Karangelos and Wehenkel, 2016a; GARPUR, 2016).

II.7 Conclusion

This paper has proposed a classification of reliability criteria according to four characteristics. A case study has illustrated how different reliability criteria can be compared in a multi-dimensional analysis of three quantitative and two qualitative indicators for short-term reliability management. This case study shows that the largest savings of expected total cost are due to a trade-off between preventive and corrective actions. A smaller portion is due to the trade-off between preventive and curtailment actions. Limits on individual or aggregate unreliability levels decrease unreliability and inequality, but increase expected total costs when compared to a full probabilistic approach.It is up to TSOs to carry out the proposed multi-dimensional analysis for their system and up to regulators to weigh the different quantitative and qualitative indicators.

Three directions for further research are identified. First, more theoretical and applied research is needed on the intermediate steps between the deterministic N-1 criterion and probabilistic criteria. This will lead to practical points of reference for TSOs to bridge the gap between them. In particular, more studies are needed that quantify ETC and other indicators of reliability criteria (reliability, equality, data needs, etc.) over a year, instead of for specific operating conditions. Second, more data are needed such that TSOs are able to apply the multi-dimensional analysis to their own large-scale systems. Third, standardized methods should be developed to help regulators decide on optimal reliability criteria.

Acknowledgements

We are grateful for very helpful comments from Fridrik Mar Baldursson, Guido Pepermans and Frank Verboven. Work is conducted through funding by the GARPUR project of the 7th European Community Framework Programme. The work of Evelyn Heylen is funded by the Research Foundation Flanders. All remaining errors are our own.

Chapter III

Cost and Quality Regulation of a Monopolist

III.1 Introduction

When regulating monopoly firms, not only is it important to give incentives for cost efficiency, incentives for optimal quality are important as well. This is especially the case for monopolies supplying necessity goods and services like electricity, gas, water, sewage, telecommunications and transport. Inadequate quality of these goods has far-reaching effects on society, ranging from decreased comfort and economics losses to diseases and deaths. As a result, many of the monopolies supplying these goods are subject to some kind of quality incentives, like minimum quality standards and quality targets with financial penalties and rewards.

Some theoretical papers have studied separately the effect of price regulation and quality regulation on quality (Spence, 1975; Sheshinski, 1976; Besanko et al., 1987; Laffont and Tirole, 1993; Sappington, 2002; Kidokoro, 2002; Weisman, 2005), but less focus has been on the joint regulation of price and quality. In addition, these papers focus on regulating optimal quantity and quality of goods, while in practice many regulatory policies focus on reducing costs and guaranteeing quality. Therefore, this paper proposes a model to jointly study the regulation of cost-efficiency and quality. To do so, we focus on the monopolist's cost function, instead of its demand curve.

This focus on cost reduction, while neglecting the effect of price and quality on demand, is particularly applicable to regulation of monopolies supplying necessity goods, as their demand does not react strongly to price and quality changes.

Many of the results in the literature on quality regulation[1] depend on the consumer demand curve. Spence (1975) has shown that an unregulated monopolist may over- or undersupply quality depending on consumers' marginal and average valuation of quality. Sheshinski (1976) extended this analysis to monopolies facing price, quantity and quality regulation. His results of optimal and regulated quality depend on quality and quantity being complements or substitutes. Similarly, the result of Besanko et al. (1987), who study the effect of price ceilings and minimum quality standards on a monopolist supplying a range of products, depends on the willingness to pay for quality. Lewis and Sappington (1992); Laffont and Tirole (1993); Weisman (2005) and Kidokoro (2002) introduce effort into their model, but they assume that costs are separable in quality and effort. They find that a price cap decreases quality, but the results of Lewis and Sappington (1992) and Laffont and Tirole (1993) hinge on quantity and quality being complements or substitutes, while Weisman (2005) assumes these cross derivatives are small. Kidokoro (2002) additionally finds that a price cap decreases investment-related quality but increases effort-related quality.

In many countries, some of the public utility industries, particularly electricity and gas, have been restructured. In that case, because of its natural monopoly properties, only the infrastructure part of the supply chain is still a regulated monopoly. In the competitive parts of the supply chain, regulation of cost efficiency and quality is less important as the market takes care of this. In the remainder of this paper, when referring to 'the monopolist', I have in mind these regulated network monopoly parts of public utility industries.

This paper makes three contributions to the literature on quality regulation. First, we explicitly identify the effect of incentive power on quality and on cost-reducing or quality-increasing effort. Second, by introducing an

[1]For an excellent overview of this literature we refer to (Sappington, 2005), and (Ajodhia and Hakvoort, 2005) for an overview focused on quality regulation of electricity distribution networks.

incentive power for both cost and quality, we are able to characterize the effect of both incentives on effort and quality. Third, we analyze the role of a monopolist's cost function that is not separable in quality and effort.

Our regulatory scheme is a linear hybrid regulation with cost and quality incentives. In the spirit of Schmalensee (1989), we depart from the optimal regulation literature and focus on 'good' linear regulatory regimes in the presence of real-life constraints such as (i) the regulator's inability to tax or subsidize the monopolist, (ii) limited information, and (iii) the regulator's and the monopolist's uncertainty about future parameter values. Because of the incentive power of the hybrid regulation, we can study the continuum between pure rate-of-return (ROR) or cost-of-service regulation, and pure price-cap (PCR) or revenue-cap regulation.

This paper is organized as follows. Section III.2 introduces a model of cost and quality, both in general terms as with specific functional forms. Next, section III.3 studies the cost and quality regulation, also both in general terms as with specific functional forms. The analysis shows that both quality and effort increase with the power of the quality incentive. The effect of the power of the cost incentive is ambiguous but under reasonable assumptions it increases effort and decreases quality. Section III.4 studies incentive power under uncertainty. Because a higher incentive power increases both socially-costly profit, and cost-reducing effort and quality, we show that the power of the cost incentive and the quality incentive are optimally below one. In section III.5, the theoretical model is compared with regulation in practice by analyzing case studies in electricity, gas and water. Finally, section III.6 concludes.

III.2 A model of cost and quality

III.2.1 General model

The most important element of our model is the cost function of the monopolist. The quantity demanded is fixed. We consider a convex total cost $C(q, e)$ of service[2] that is increasing in the quality level q and decreasing in

[2]In the remainder of the paper, this cost is referred to as service cost, as most monopolies supply a service, rather than a good. For example, electricity, water, gas, telecom-

cost-reducing effort e:

$$C'_q > 0, \quad C''_{qq} \geq 0, \quad C'_e < 0, \quad C''_{ee} \geq 0 \quad \text{and} \quad C''_{qq}C''_{ee} - (C''_{qe})^2 \geq 0 \quad (3.1)$$

That is, the marginal cost increase of quality increases with quality and the marginal cost decrease of effort decreases with effort.

The effort e can be interpreted as all monetary and non-monetary actions that decrease the cost of supplying the monopolist's service or good at a certain quality level q. If effort increases, a monopolist can decrease costs without decreasing quality or similarly, increase quality without increasing cost. Examples of cost-reducing effort are investments in lower power losses in electricity networks or in better maintenance technologies for gas pipelines. Equivalently, examples of quality-increasing effort are cutting trees near electricity lines, controlling the pressure of gas pipelines, checking water quality, increasing the number of airport security guards, etc. Exerting this effort e entails a cost or disutility $\psi(e)$ for the monopolist. This cost is increasing and convex in effort. As in (Schmalensee, 1989), notation and prose are simplified by treating $\psi(e)$ as a pecuniary cost.

Consumers derive a total benefit $V(q)$ from consuming the good or service with quality level q. This benefit is increasing and concave in the quality level. To facilitate the comparative statics exercise, suppose that $V(q) = \overline{V}v(q)$, where $\overline{V} > 0$ is the valuation of quality. Estimates of $\overline{V}$ exist, for example in electricity, where it is called the value of lost load (VOLL). It represents the cost of unserved electricity and is expressed per quantity demanded (e.g. €/MWh) (Ovaere et al., 2016).

Lastly, to ensure the monopolist a non-zero profit, he should be paid at least the sum of effort and service costs. Since regulators are (generally) not able to tax or subsidize the monopolist, we assume that these costs are completely covered by consumer payments.[3] For example, network tariffs for electricity transmission and distribution networks, capacity and commodity charges for gas, access charges for telecommunication services, infrastruc-

munication services, transport, etc.

[3]Pricing of a natural monopoly is an important question but is not dealt with in this paper, as we assume demand to be inelastic to changes in price and quality. In practice average cost pricing and uniformly increasing prices above marginal costs are used (Laffont, 1994), while Ramsey-Boiteux pricing is the theoretical optimum (Boiteux, 1956).

ture charges for railroads, tolls for toll roads, airport fees, etc. In most network industries the monopolist is not free to choose the price structure independently. Only in the telecommunications industry, companies have some freedom to choose prices. The sum of all consumer payments should be sufficient to guarantee the monopolist at least a non-negative profit in all possible states of nature to guarantee service, especially when the product or service sold is a necessity (such as water, gas or electricity) (Schmalensee, 1989). As a result, the profit of the monopolist is:

$$\Pi = \text{consumer payments } - C(e,q) - \psi(e) \geq 0 \tag{3.2}$$

We suppose that the regulator strictly prefers consumer surplus (CS) to monopoly profit, i.e. social welfare = CS + $\alpha\Pi$, with $\alpha < 1$ (Baron and Myerson, 1982; Armstrong and Sappington, 2004). Therefore, it is optimal to just repay the monopolist's network and effort costs, and keep socially costly monopoly profit Π equal to zero. This is only possible under perfect information. Under imperfect information, the monopolist earns a positive expected profit (see section III.4).

Combining the above assumptions, a welfare-maximizing regulator chooses effort e and quality q to maximize the following function:

$$\max_{\{q,e\}} \quad V(q) - \Big(C(q,e) + \Pi + \psi(e)\Big) + \alpha\Pi \tag{3.3}$$

$$= V(q) - \Big(C(q,e) + \psi(e)\Big) - (1-\alpha)\Pi \tag{3.4}$$

$$\text{s.t. } \Pi \geq 0 \tag{3.5}$$

This leads to the following first-order conditions for optimal quality and effort:

$$\begin{aligned} V'_q - C'_q &= 0 \\ C'_e + \psi'_e &= 0 \end{aligned} \tag{3.6}$$

The first-order conditions show that the quality level is optimal at the point where the marginal increase of consumer benefit equals the marginal social cost of the cost increase; and the effort is optimal at the point where the marginal decrease of service costs equals the marginal increase of effort costs. Since effort and service costs are convex and consumer benefit is concave, a global maximum exists.

Totally differentiating the first-order conditions, we obtain the following proposition:

Proposition 1. *If total cost is convex (assumptions* (3.1)*), optimal quality increases with the valuation of quality:*

$$\frac{dq^*}{d\overline{V}} \geq 0$$

The effect of the valuation of quality on optimal cost-reducing effort depends on the cross derivative of service costs C''_{eq}:

$$\frac{de^*}{d\overline{V}} \geq 0 \quad \textit{if} \quad C''_{eq} \leq 0 \qquad \textit{and} \qquad \frac{de^*}{d\overline{V}} \leq 0 \quad \textit{if} \quad C''_{eq} \geq 0$$

Proof. The total derivatives of the first-order conditions are:

$$\begin{bmatrix} C''_{qq} - V''_{qq} & C''_{qe} \\ C''_{eq} & \psi''_{ee} + C''_{ee} \end{bmatrix} \begin{bmatrix} dq^* \\ de^* \end{bmatrix} = \begin{bmatrix} v'_q \\ 0 \end{bmatrix} d\overline{V} \tag{3.7}$$

which leads to the following expressions:

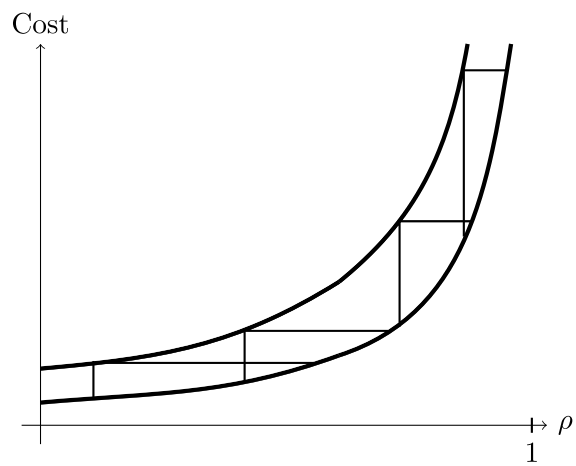
$$\frac{dq^*}{d\overline{V}} = \frac{v'_q(\psi''_{ee} + C''_{ee})}{(C''_{qq} - V''_{qq})(\psi''_{ee} + C''_{ee}) - C''^2_{qe}} = \frac{\geq 0}{\geq 0} \tag{3.8}$$

$$\frac{de^*}{d\overline{V}} = \frac{-v'_q C''_{eq}}{(C''_{qq} - V''_{qq})(\psi''_{ee} + C''_{ee}) - C''^2_{qe}} = \frac{\geq \ \text{or} \ \leq 0}{\geq 0} \tag{3.9}$$

□

The first part of this proposition is straightforward. The optimal quality level increases if quality is valued more. The second part, however, depends on the cross derivative of service costs. If it is negative, effort increases with the valuation quality. Figure III.1 shows two service cost functions with negative cross derivative. The upper curve is with low effort, the lower curve with higher effort. A negative cross derivative means that the marginal cost decrease of cost-reducing effort increases with the level of service costs (vertical lines), or alternatively, the marginal quality increase of quality-increasing effort decreases with the level of service costs (horizontal lines). On the other hand, if the cross derivative of service costs is positive, effort decreases with the valuation of quality. However, the cross derivative

Figure III.1: Cost-reducing effort (vertical lines) and quality-increasing effort (horizontal lines) for a negative cross derivative of service costs.



is unlikely to be positive in reality, as this would mean that the marginal quality increase of quality-increasing effort increases with the level of service costs.

To illustrate the above analysis, the next section proposes functional forms for network and effort costs. The service cost has a negative cross derivative in this illustration.

III.2.2 Model with specific functional forms

First, we consider the following service cost function, which satisfies assumptions (3.1):

$$C(q,e) = \frac{\beta}{e(1-q)}, \quad q \in [0,1], \quad \beta, e > 0 \tag{3.10}$$

where the cost parameter β determines the service cost for a given level of effort e and quality q. The cost parameter β is exogenous and depends, inter alia, on consumer density, geography, climate, consumer density, employee wages, yearly demand, size of the network, economies of scale, and economies of scope. Quality varies between zero and one, with a value of one indicating a perfect quality level and zero indicating no quality. The service costs

$C(q,e)$ increase with quality.[4] The cost of reaching a perfect quality level ($q = 1$) is infinite. For example, a completely reliable gas or electricity network requires large reliability margins and thus many redundant lines and pipelines (Ovaere and Proost, 2016). Similarly, to completely eliminate congestion, an airport needs many runways, a railway network many tracks, a toll road many parallel lanes, and a telecommunications network many links and towers.

Second, effort costs have the following functional form:

$$\psi(e) = \frac{E}{2}e^2, \quad E > 0 \tag{3.11}$$

where the effort cost parameter E is strictly positive. A higher effort leads to a higher quality level for the same service costs or lower service costs for the same quality level. If effort is higher than one, service costs decrease; if effort is lower than one, service costs increase. Lastly, consumer benefit is linear in quality: $V(q) = \overline{V}q$.

Putting together the above functional forms, the optimal quality and effort are:[5]

$$\begin{aligned} q^* &= 1 - \left(\frac{E\beta^2}{\overline{V}^3}\right)^{0.2} \\ e^* &= \left(\frac{\overline{V}\beta}{E^2}\right)^{0.2} \end{aligned} \tag{3.12}$$

These expressions show that the optimal quality decreases with effort cost and service cost, and – as predicted by proposition 1 – increases with V. The optimal effort level decreases with effort cost, increases with service cost and – again, as predicted by proposition 1 – increases with V.

As an illustration for electricity networks, take the following numerical values: $E = 0.4, \beta = 0.000078$ and $\overline{V}$=VOLL=5000 [€/MWh]. This yields an optimal effort of e^*=1.05, which means a cost decrease of about 5%. The

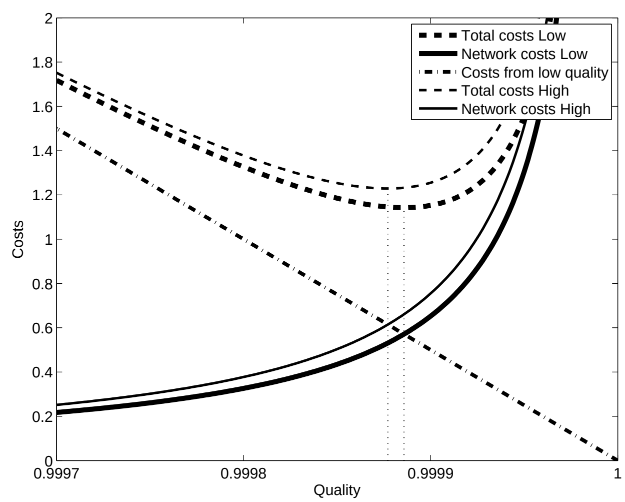
[4]Reichl et al. (2008) confirm that the reliability level of transmission networks increases with service costs. They find that annual average interruption duration increases by 1.36 minutes if average service costs decrease by 1€/MWh. Likewise, in a case study of Italian distribution system operators (DSOs), Cambini et al. (2016) find that operating expenditures and capital expenditures affect service quality.

[5]Appendix III.A repeats the analysis of this section for different functional forms and finds similar results.

optimal quality level is q^*=0.999885, which means a reliability of electricity supply of 99.9885%, or equivalently an unavailability of 60 minutes per year.

If, ceteris paribus, the cost β is 20% higher, the optimal effort is e^*_H=1.09 and the optimal quality level is q^*_H=0.999877. That is, if the monopolist is high cost (a higher β), it is optimal to exert more cost-reducing effort and have a lower quality level. Figure III.2 exemplifies the above illustration. This figure plots the convex service costs (solid lines) and the total costs (dotted lines) for a low value of β (thick lines) and a high value of β (thin lines), and the lost benefit from a less than perfect quality (dash-dot lines), i.e. the cost $\overline{V}(1-q)$.

Figure III.2: Optimal quality level as a function of service costs (solid lines) and the costs of low quality (dash-dot lines), with A = 0.4, V = 5000, β_L = 0.000053 and β_H = 1.2β_L.



This section dealt with optimal cost and quality of a monopolist. The next section analyses the cost and quality regulation of a monopolist.

III.3 Cost and quality regulation

As Holmstrom and Milgrom (1987) and Schmalensee (1989) have noted, most incentive schemes observed in practice are linear. In addition, Holmstrom and Milgrom (1987) note that linear contracts enjoy a robustness that makes them effective in a wide range of situations. Therefore, as in (Schmalensee, 1989), we propose a linear regulatory scheme that leads to 'good' results in many settings, but is not optimal for all functional forms and distributions of uncertainty. Contrary to Schmalensee (1989), who assumes simple functional forms and does not arrive at analytical solutions, we arrive at insightful comparative static results without functional forms and analytical solutions with a fairly complex functional form for service costs. The main difference with Schmalensee is that we assume that demand is perfectly inelastic. In this section, we first study a general model of cost and quality regulation, then we discuss the model with specific functional forms.

III.3.1 General model of cost and quality regulation

The previous section studied optimal quality and effort of a monopolist. In reality, however, the monopolist's objective function can diverge from welfare maximization. Therefore, a regulator, who is assumed to be a benevolent maximizer of social welfare,[6] designs a regulatory scheme that attempts to give the monopolist the correct incentives to align its profit maximization objective with welfare maximization. This paper studies a revenue cap R with cost-efficiency and quality incentives:

$$R = b\overline{C} + (1-b)C + b_V\overline{V}(q-\overline{q}) \text{ with } b \in [0,1] \tag{3.13}$$

Each year the regulated monopolist earns a revenue R, which depends on a cost norm $\overline{C}$, its realized service costs C, a quality norm $\overline{q}$, and its realized quality level q. The allowed revenue R increases with both realized service costs and realized quality.[7] However, the regulated monopolist is remunerated only a fraction (1-b) of its realized service costs and earns the remaining fraction b of the ex-ante determined cost norm $\overline{C}$, also called justified costs. This cost norm is determined by the regulator based on the regulated

[6] A regulatory agencies can also be captured by interest groups, as analyzed in Stigler (1971) and Laffont and Tirole (1993, chapter 11)

[7] We assume that quality is verifiable, such that a regulator can use observed quality, without needing to resort to sales incentives and threats to reputation (Laffont and Tirole, 1991; Lewis and Sappington, 1991).

monopolist's historical costs, on benchmark analyses of other monopolists (Schleifer, 1985)[8], and on assumptions of the growth rate of the retail price minus the anticipated rate of technological progress (RPI-X) (Littlechild, 1983). The fraction b is the power of the cost incentive.[9] The higher b, the more of its revenue is allocated independently of realized costs; the lower b, the more of its service costs are remunerated in function of realized costs. As before, revenue is raised through consumer payments.[10]

A regulatory scheme with $b = 1$ amounts to a pure revenue cap, while $b = 0$ amounts to cost-of-service regulation. There is a sizable literature that studies the properties, advantages and disadvantages of these two limits of the regulatory spectrum. For an excellent overview we refer to (Liston, 1993; Laffont and Tirole, 1993) and (Decker, 2014).

Revenue cap (3.13) is sometimes expressed in terms of a cost-sharing parameter, for example in the regulation of electricity networks in Great Britain (see section III.5.3.1). The cost-sharing parameter s is the fraction of cost under- or overperformance that is paid by consumers. In that case the cost incentive of the revenue cap is:

$$R = \overline{C} + s(C - \overline{C}) = (1 - s)\overline{C} + sC \tag{3.14}$$

which shows that the cost-sharing parameters is one minus the incentive power, i.e. $s = 1 - b$.

To revenue cap (3.13), a quality incentive is added. If quality q is above the quality norm $\overline{q}$, the monopolist earns an additional revenue of $b_V\overline{V}(q-\overline{q})$; if the quality is lower than the quality norm, a penalty of $b_V\overline{V}(q - \overline{q})$ is subtracted from its allowed revenue. The parameter b_V is the power of the quality incentive. Similar to the power b of the cost incentive, a high power

[8]See (Estache et al., 2004) and (Giannakis et al., 2005) for examples of electricity network benchmarking.

[9]We assume that the rate of return on realized service costs $(1 - b)C$ is equal to the weighted average cost of capital (WACC) and thus does not end up in the profit of the monopolist. If the rate of return is r [%] higher than the WACC, the analysis is still valid but the power of the cost incentive is adjusted downwards: $b_r = b - (r - WACC)(1 - b)$.

[10]Since consumer charges are often set at the beginning of a year or regulatory period, and total consumption and realized service costs are only known at the end, an excess or deficit of raised revenue is possible. However, consumer charges are set such that the excess or deficit revenue balance is zero over time.

of the quality incentive means that deviations from the quality norm have a larger effect on the monopolist's allowed revenue. The parameter $\overline{V}$ is the (average) quality valuation of consumers.

To summarize, the monopolist's allowed revenue depends on ex-ante determined cost and quality norms, and on ex-post realized service cost and quality. However, cost and quality norms could depend on ex-post parameters, such as peak or total demand, to eliminate demand effects.

If a monopolist is subject to revenue cap (3.13), its profit function to maximize is:

$$\max_{\{e_M, q_M\}} b(\overline{C} - C(q_M, e_M)) + b_V \overline{V}(q_M - \overline{q}) - \psi(e_M) \tag{3.15}$$

where e_M and q_M are the effort and quality chosen by the monopolist. This leads to the following first-order conditions for the monopolist's choice of quality and effort:

$$\begin{aligned} b_V \overline{V} - bC'_q &= 0 \\ bC'_e + \psi'_e &= 0 \end{aligned} \tag{3.16}$$

The first-order conditions tell that the monopolist chooses the quality level such that the marginal increase of its regulated revenue from the quality incentive equals the marginal increase of its regulated revenue from the cost incentive; and it chooses its effort such that the marginal increase of its regulated revenue from the cost incentive equals its marginal cost of effort. As before, a global maximum exists because effort and service costs are convex and the quality incentive is concave.

Under perfect information the cost norm $\bar{C}$ and quality norm $\overline{q}$ can be chosen such that the socially-costly profit of the monopolist is zero.

Comparing first-order conditions (3.6) and (3.16) leads to a straightforward proposition:

Proposition 2. *If a monopolist is regulated according to revenue cap* (3.13) *and the regulator has perfect information, its quality and effort are optimal at incentive powers* $b = 1$ *and* $b_V = 1$.

This result is in line with the optimal regulation literature (Laffont and Tirole, 1986, 1993) that finds that incentive power equals one under perfect information[11], while under imperfect information a low-cost (efficient) monopolist self-selects a higher incentive power if a menu of linear contracts is offered.

In addition to proposition 2, total differentiation of the monopolist's first-order conditions leads to the following proposition:

Proposition 3. *If total cost is convex (assumptions* (3.1)*), the power of the quality incentive b_V increases the monopolist's quality and has an ambiguous effect on the monopolist's effort:*

$$\begin{aligned} &\frac{dq_M}{db_V} \geq 0 \\ &\frac{de_M}{db_V} > 0 \quad \textit{if} \quad C''_{eq} < 0 \qquad \textit{and} \qquad \frac{de_M}{db_V} \leq 0 \quad \textit{if} \quad C''_{eq} \geq 0 \end{aligned}$$

The power of the cost incentive b has an ambiguous effect on both the monopolist's quality and effort:

$$\begin{aligned} &\frac{dq_M}{db} \geq \textit{or} \leq 0 \quad \textit{if} \quad C''_{eq} \leq 0 \qquad \textit{and} \qquad \frac{dq_M}{db} < 0 \quad \textit{if} \quad C''_{eq} > 0 \\ &\frac{de_M}{db} \geq \textit{or} \leq 0 \quad \textit{if} \quad C''_{eq} \leq 0 \qquad \textit{and} \qquad \frac{de_M}{db} > 0 \quad \textit{if} \quad C''_{eq} > 0 \end{aligned}$$

Proof. The total derivatives of the first-order conditions are:

$$\begin{bmatrix} bC''_{qq} & bC''_{qe} \\ bC''_{eq} & \psi''_{ee} + bC''_{ee} \end{bmatrix} \begin{bmatrix} dq_M \\ de_M \end{bmatrix} = \begin{bmatrix} \overline{V} \\ 0 \end{bmatrix} db_V \tag{3.17}$$

$$\begin{bmatrix} bC''_{qq} & bC''_{qe} \\ bC''_{eq} & \psi''_{ee} + bC''_{ee} \end{bmatrix} \begin{bmatrix} dq_M \\ de_M \end{bmatrix} = \begin{bmatrix} -C'_q \\ -C'_e \end{bmatrix} db \tag{3.18}$$

[11]To our knowledge the optimal regulation literature does not say anything about the incentive power of the quality incentive.

which leads to the following expressions:

$$
\begin{aligned}
\frac{dq_M}{db_V} &= \frac{\overline{V}(\psi''_{ee} + bC''_{ee})}{bC''_{qq}(\psi''_{ee} + bC''_{ee}) - b^2 C''^2_{qe}} = \frac{\geq 0}{\geq 0} \\
\frac{de_M}{db_V} &= \frac{-bC''_{eq}}{bC''_{qq}(\psi''_{ee} + bC''_{ee}) - b^2 C''^2_{qe}} = \frac{\geq or \leq 0}{\geq 0} \\
\frac{dq_M}{db} &= \frac{-C'_q(bC''_{ee} + \psi''_{ee}) + bC'_e C''_{eq}}{bC''_{qq}(\psi''_{ee} + bC''_{ee}) - b^2 C''^2_{qe}} = \frac{\geq or \leq 0}{\geq 0} \\
\frac{de_M}{db} &= \frac{-bC''_{qq}C'_e + bC''_{eq}C'_q}{bC''_{qq}(\psi''_{ee} + bC''_{ee}) - b^2 C''^2_{qe}} = \frac{\geq or \leq 0}{\geq 0}
\end{aligned}
\tag{3.19}
$$

□

The results of proposition 2 depend on the cross derivative of service costs. If this is zero (e.g. $C = q - e$, as in (Laffont and Tirole, 1993, Chapter 4)), quality decreases with the power of the cost incentive and increases with the power of the quality incentive, while effort decreases with the power of the cost incentives and does not change with the power of the quality incentive. On the other hand, if the cross derivative is negative, a higher power of the quality incentive increases both quality and effort. The effect of the power of the cost incentive is ambiguous. Both quality and effort could increase or decrease with incentive power.

Proposition 3 predicts the effect of the power of the cost and the power of the quality incentive on both quality and effort. The theoretical literature focused so far on the effect of the cost incentive on quality. Their prediction is in line with proposition 3. Sheshinski (1976) finds that with a positive cross derivative of inverse demand with respect to quality and quantity ($p_{qx} > 0$), a price cap decreases quality if the cross derivative of profits is negative ($\pi_{qx} < 0$) and increases quality if the cross derivative of profits is positive ($\pi_{qx} > 0$). The effect is ambiguous for a negative cross derivative of inverse demand. Similarly, Kidokoro (2002) finds that quality decreases with incentive power for investment-related quality and increases with incentive power for effort-related quality.

The theoretical predictions of proposition 3 are confirmed by the (small) empirical literature of quality in electricity networks. Ter-Martirosyan and Kwoka (2010) find that incentive regulation is associated with significantly longer duration of service outages ($\frac{dq}{db} < 0$), but that this quality reduction is offset when regulation includes quality incentives ($\frac{dq}{db_V} > 0$). Similarly, Schmidthaler et al. (2015) find that the introduction of quality incentives leads to reductions of the annual outage duration by 16.05% on average ($\frac{dq}{db_V} > 0$). In the telecommunications industry, the empirical results (in the U.S.) have been mixed (Banerjee, 2003; Sappington, 2003; Ai et al., 2003). However, since cost and quality incentives were introduced simultaneously in many states, their effects are difficult to disentangle ($\frac{dq}{db_V db} >< 0$).

III.3.2 Model of cost and quality regulation with specific functional forms

Using the functional forms of section III.2.2, the monopolist's quality and effort are:

$$\begin{aligned} q_M &= 1 - \left(\frac{b^2 E \beta^2}{b_v^3 \overline{V}^3}\right)^{0.2} \\ e_M &= \left(\frac{b_V b \overline{V} \beta}{E^2}\right)^{0.2} \end{aligned} \tag{3.20}$$

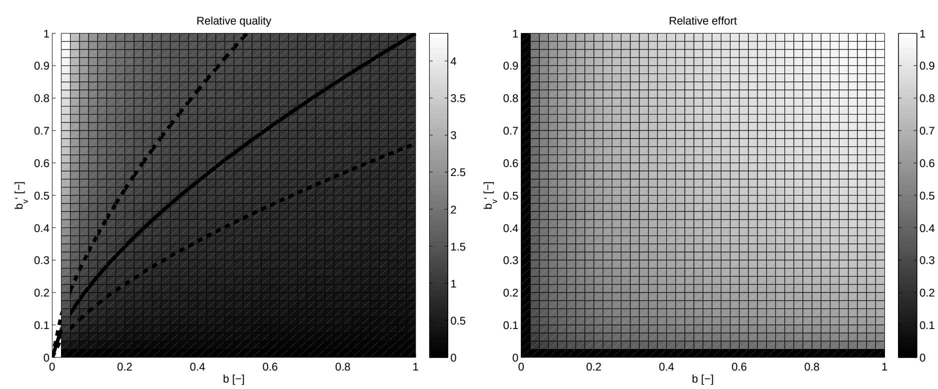
These expressions show that both quality and effort increase with the power b_V of the quality incentive – as predicted by proposition 3, because $C''_{qe} < 0$ for our assumed cost function. Proposition 3 is ambiguous on the effect of b on quality and effort, but the above expressions show that for our assumed cost function, quality decreases and effort increases with the power b of the cost incentive. The alternative functional forms in Appendix III.A yield similar results.

Figure III.3 plots the monopolist's relative effort and quality as a function of the power of the cost incentive (b) and the quality incentive (b_V).[12] Relative effort is defined as the ratio of the monopolist's effort and the optimal effort, while relative quality is defined as the inverse ratio of unreliability (1-quality). For example, a relative quality of 2 means that the monopolist's unreliability (e.g. expressed in interrupted or congested hours) is half the optimal unreliability. Darker regions represent higher relative values. The

[12] It should be noted that the values depend on functional form assumptions but not on parameter values.

left graph of Figure III.3 shows that the monopolist's quality is too high in the upper-left corner, where the quality incentive is high and the cost incentive is low. Similarly, its quality is too low in the lower-right corner, where the quality incentive is low and the cost incentive is high. The black line in the center indicates the incentive values that lead to the optimal quality, the dashed lines indicate a relative of 2 and 0.5 respectively. Evidently, this line goes through $b = b_V = 1$. The right graph of Figure III.3 shows that the monopolist's effort increases with both the quality and the cost incentive. In the limit of a strict cost-of-service regime ($b = 0$) the effort is zero and the quality is infinitely high if the quality incentive is positive. In the limit of a strict price cap regime ($b = 1$) the quality decreases with a decreasing quality incentive. Of course, the theoretical result of both limits exaggerates the monopolist's behavior in practice. Quality and cost are bounded by opposition from both the regulator and consumers to unacceptably high or low levels (Sappington and Weisman, 2010).

Figure III.3: Relative quality $\frac{1-q^*}{1-q_M}$ and relative effort $\frac{e_M}{e^*}$ for different values of b and b_V



III.4 Cost and quality regulation under uncertainty

Proposition 2 stated that under perfect information both the power of the cost and the quality should equal one, because the cost norm $\bar{C}$ and quality norm $\overline{q}$ can be chosen such that the socially-costly profit of the monopolist is zero. In reality, however, the regulator is uncertain about current costs and future cost realizations. As the regulator should guarantee the monopolist a non-negative profit in all possible states of nature, he can at best choose the norms such that the monopolist's profit is zero in the worst-case scenario. In other states of nature the monopolist's profit will be positive and his expected profit E[Π] will thus be positive. The regulator's maximization problem is:

$$\begin{aligned} \max_{\{b,b_V\}} \quad & V(q_M) - \Big(C(q_M,e_M) + \psi(e_M)\Big) - (1-\alpha)\mathrm{E}\big[\Pi(q_M,e_M,\epsilon)\big] \\ \text{s.t.} \quad & \min_{\{\epsilon\}} \Pi(q_M,e_M,\epsilon) \geq 0 \end{aligned} \tag{3.21}$$

where $q_M(b,b_V)$ and $e_M(b,b_V)$ are determined from equation (3.16). The parameter ϵ represents the uncertainty about the monopolist's parameters. A higher ϵ means higher uncertainty (Schmalensee, 1989). Asymmetric information is not assessed. We assume that the monopolist does not have more information about future parameter values than the regulator or that he does not use this asymmetric information to his benefit. The constraint ensures that the monopolist's profit is at least non-negative for all values of the parameter ϵ. We assume the functional forms of section III.2.2 and assume that the regulator knows these functional forms but is uncertain about the values of the parameters E, β and V. To derive meaningful analytical results, we assume the functional forms of section III.2.2. Suppose that his aggregate uncertainty about these parameters is characterized by $\left(E\beta\overline{V}^2\right)^{0.2}(1+\epsilon)$, where the regulator knows that ϵ has a symmetric probability density function f(ϵ) with support [-$\bar{\epsilon}$, $\bar{\epsilon}$]. This leads to the following lemma:

Lemma 1. *Under the above assumptions, the fixed payment and the expected profit increase with the power of both the cost and quality incentive.*

Proof. Using expressions (3.20) of the monopolist's choice of effort and quality, expressions for service costs, effort costs and the cost of unreliability (1-q) are:

$$C(q,e) = A\frac{b_V^{0.4}}{b^{0.6}},\ \psi(e) = \frac{A}{2}b_V^{0.4}b^{0.4} \text{ and } (1-q)V = A\frac{b^{0.4}}{b_V^{0.6}}, \text{ with } A = \left(E\beta^2\overline{V}^2\right)^{0.2} \tag{3.22}$$

To have $\Pi(\epsilon = \overline{\epsilon}) = 0$, fixed payments should equal:

$$b\overline{C} + b_V\overline{V}(1-\overline{q}) = 2.5Ab_V^{0.4}b^{0.4}(1+\overline{\epsilon}) \tag{3.23}$$

The expected profit is:

$$\mathrm{E}\big[\Pi(q_M, e_M)\big] = 2.5Ab_V^{0.4}b^{0.4}\int_{-\overline{\epsilon}}^{\overline{\epsilon}}(\overline{\epsilon}-\epsilon)f(\epsilon)\mathrm{d}\epsilon = 2.5Ab_V^{0.4}b^{0.4}\overline{\epsilon} \tag{3.24}$$

because $\int_{-\overline{\epsilon}}^{\overline{\epsilon}}(\overline{\epsilon}-\epsilon)f(\epsilon)\mathrm{d}\epsilon = \overline{\epsilon}$ for symmetric distributions. □

This result is in line with the simulation model of Gasmi et al. (1994) and empirical evidence in the telecommunications industry (Sappington, 2002; Hauge and Sappington, 2010). Since this lemma shows that expected profit increases with the power of both the cost and quality incentives, the regulator faces a trade-off between optimal cost and quality incentives, and higher expected profit for the monopolist. Increasing the incentive powers increases socially-costly profit but increases cost-reducing effort and quality (see proposition 3). This results in the following proposition:

Proposition 4. *In the presence of uncertainty ($\overline{\epsilon} > 0$) and socially-costly consumer payments, and assuming the above functional forms, the optimal power of the cost incentive and the quality incentive are equal and below one:*

$$b^* = b_V^* = \frac{1}{1+5(1-\alpha)\overline{\epsilon}} < 1 \tag{3.25}$$

Proof. Inserting the functional forms and the expression for expected profit from Lemma 1 into the regulator's maximization (3.21) yields the following expression:

$$\max_{\{b,b_V\}} V - A\frac{b^{0.4}}{b_V^{0.6}} - \left(A\frac{b_V^{0.4}}{b^{0.6}} + \frac{A}{2}b_V^{0.4}b^{0.4}\right) - 2.5(1-\alpha)Ab_V^{0.4}b^{0.4}\overline{\epsilon} \tag{3.26}$$

Which leads to the following first-order conditions:

$$2\frac{b}{b_V} - 3 + b + 5(1-\alpha)\overline{\epsilon}b = 0 \tag{3.27}$$

$$-3 + 2\frac{b_V}{b} + b_V + 5(1-\alpha)\overline{\epsilon}b_V = 0 \tag{3.28}$$

which results in $b^* = b_V^* = \frac{1}{1+5(1-\alpha)\overline{\epsilon}}$. Appendix III.A shows that the result is robust for different functional forms. □

This proposition shows that the power of the cost and quality incentives (i) should be equal, and (ii) should decrease with increasing uncertainty and decreasing preference for monopoly profit. This last result is similar to Schmalensee (1989) who states that ROR regulation is preferred at high levels of uncertainty. However, in our case the incentive power decreases with the level of uncertainty and only equals ROR regulation in the limit. Comparing with expressions (3.20), both effort and quality are thus lower under uncertainty than under perfect information.[13] Section III.5.3 will show case studies for which the incentive power of the cost incentive is between 0.46 and 0.63. This value is in line with the prescription of equation (3.25). For example, if $\alpha = 0.5$ and $\bar{\epsilon} = 0.4$, i.e. monopoly profit is valued 50% less than consumer surplus and the monopolist has an uncertainty of [-40%,+40%] over the parameter values, optimal incentive power equals 0.5.

From proposition 4 it is straightforward to show the following corollary:

Corollary 1. *Under the above assumptions, expected total cost of the monopolist increases with the power of the quality incentive, and decreases with the power of the cost incentive, except for very high uncertainty ($\bar{\epsilon}$) and a very low preference for monopoly profit (α).*

Proof.

$$\begin{aligned} TC &= C(q_M, e_M) + \psi(e_M) + (1-\alpha)\mathrm{E}\big[\Pi(q_M, e_M)\big] \\ &= Ab_V^{0.4}\Big(\frac{1}{b^{0.6}} + \frac{b^{0.4}}{2} + 2.5(1-\alpha)\bar{\epsilon}b^{0.4}\Big) \end{aligned} \tag{3.29}$$

□

When the parameters E, β and V are expressed on a per quantity basis, the above result also applies to the average price or consumer tariff. The theoretical prediction of corollary 1 is also confirmed by the empirical literature of price-cap regulation in the U.S. telecommunications industry. After introduction of price-cap regulation, Mathios and Rogers (1989) find significantly lower rates, Kaestner and Kahn (1990) find lower prices, Ai

[13] A related setting that leads to similar conclusions about the optimal incentive power is one where the cost and quality norms are determined endogenously according to some backward-looking rule, see appendix III.B.

and Sappington (2002) find that costs are generally lower, and Blank et al. (1998) find no evidence of reduced prices.

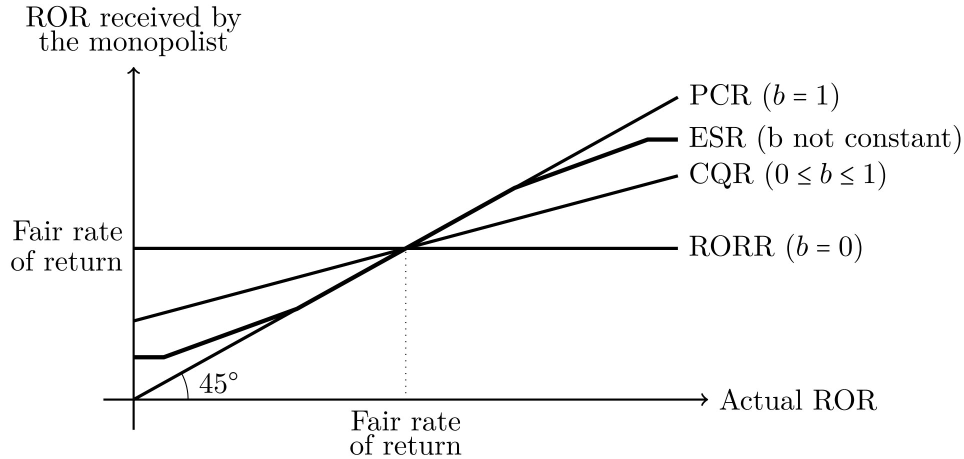
III.5 Regulation in practice

III.5.1 The difference between different regulations

Figure III.4 compares the the cost and quality regulation of this paper (CQR) with rate-of-return regulation (RORR), price-cap regulation (PCR) and earning-sharing regulation (ESR) (Schmalensee, 1989; Gasmi et al., 1994; Lyon, 1996; Weisman, 2005; Blank and Mayo, 2009). The figure links the actual rate of return with the rate of return received by the monopolist. Under RORR the rate of return received by the monopolist is always the same, irrespective of its cost-reducing efforts or exogenous cost shocks. On the other side of the spectrum, under PCR the monopolist receives all changes of the actual rate of return, irrespective of being due to exogenous factors or endogenous cost-reducing effort by the monopolist. The cost and quality regulation (CQR) of this paper is in between these two extremes. The monopolist receives only fraction b of changes of the actual rate of return. This linear regulation is a simple form of earning-sharing regulation (ESR), or equivalent profit-sharing or sliding-scale regulation, which has a non-constant power of the cost incentive and caps on minimum and maximum received rate of return. As this allows for more degrees of freedom than the linear model of this paper, ESR probably leads to a higher efficiency, if designed properly.[14] Therefore, having a more intricate regulation is a trade-off between providing better incentives to the monopolist and the cost of needed information. As an example, in their search for explanations for the short tenure of ESR in the US telecommunications industry, Sappington and Weisman (2010, p.246) state that ESR introduces contentious technical issues because the sharing rule depends on the level of measured earnings. This is not a problem for linear regulation as the incentive power is constant.

[14] However, since analytical expressions are not possible anymore, simulations are needed to assess if ESR is more optimal than our analyzed CQR, and which extensions lead to the largest welfare gain: caps, asymmetry and the number of steps in the incentive power. This is an interesting avenue for future research, a point also noted by (Joskow, 2014).

Figure III.4: Comparing rate-of-return regulation (RORR), earning-sharing regulation (ESR), the cost and quality regulation of this paper (CQR) and price-cap regulation (PCR).



III.5.2 Scope of the revenue cap

Up to now, we assumed that all costs of a monopolist are included under the revenue cap. However, if only a part of a monopolist's total costs are subject to cost incentives and his other costs earn a regulated rate of return, the monopolist may distort the relative use of its costs, similar to the famous Averch-Johnson effect (Averch and Johnson, 1962).[15] For example, in the regulation of transmission networks in Sweden, cost incentives only apply to operating costs (OC), while capital costs (CC) earn a regulated rate of return (NordREG, 2012).

If operating costs are decreasing in capital costs, e.g. $OC = \frac{\beta/CC}{e(1-q)}$, monopolists with cost incentives on operating costs and a regulated rate of return on capital costs have a tendency to overinvest in capital costs to decrease operating costs. If a quality incentive is included, the monopolist's

[15] Giannakis et al. (2005) find a trade-off between operating and capital expenditures in UK electricity distribution networks.

effort and quality will be (similar to equations (3.20)):

$$\begin{aligned} e_M &= (\frac{bb_V\overline{V}\beta}{E^2CC})^{0.2} \\ q_M &= 1-(\frac{b^2E\beta^2}{b_V^3\overline{V}^3CC^2})^{0.2} \end{aligned} \tag{3.30}$$

In theory this leads to zero effort and perfect quality, but in reality the decrease of operating costs and the increase of capital costs are limited by regulatory, societal and political constraints.

As a reaction to this distortion of costs, OFWAT, the Water Services Regulation Authority of England and Wales, changed its regulation from one where operating and capital costs were regulated separately (2010-2015), to one where all costs are regulated under the same revenue cap (2015-2020), a so-called Totex approach, see section III.5.3.3.

III.5.3 Case studies

This sections analyzes case studies in electricity, gas and water networks, where hybrid revenue caps are used in combination with quality incentives. We will see that incentive power of the quality incentive is equal to one in all case studies, except for regulation of electricity networks in Norway, where it is equal to 0.6. On the other hand, the incentive power of the cost incentive is between 0.46 and 0.63 in all case studies.

III.5.3.1 Regulation of electricity transmission and distribution networks in Norway and Great Britain

Electricity transmission and distribution networks in Norway are regulated using the cost and quality revenue cap regulation that is studied in section III.3 of this paper. In its most simplified form, the Norwegian regulation is:

$$R = 0.6\overline{C} + 0.4C + 0.6\overline{V}(q-\overline{q}) \tag{3.31}$$

That is, the power of the cost and quality incentives are equal and equal to 0.6. The cost and quality norms are based on historical values and on benchmarking analysis.[16] The quality incentive adds or subtracts an amount from the allowed revenue cap of the transmission or distribution system operator, so that the quality improvements or deteriorations are socialized over all consumers. There is no direct compensation to affected consumers. However, the treatment of quality incentives is much more elaborate than presented in equation (3.31). The quality norm is formulated on the basis of interruption costs, a method known as the cost of energy not supplied (CENS) regulation. In the CENS regulation, interruption costs IC are calculated for different consumer groups c, and both the time t and duration d of interruptions u [MWh] have an effect on interruption costs (Kjolle et al., 2008). In summary:

$$R = 0.6\overline{C} + 0.4C + 0.6(\overline{IC} - \sum_{c,t,d} \overline{V}(c,t,d)u(c,t,d)) \tag{3.32}$$

Similar regulation exists in many other countries such as England, Wales, Scotland, the Netherlands (Hesseling and Sari, 2006) and Australia[17]. In Great Britain, the three transmission owners – National Grid Electricity Transmission (NGET TO) in England and Wales, Scottish Hydro Electric Transmission (SHE) in the north of Scotland, and SP Transmission (SPT) in the south of Scotland– and the system operator (NGET SO) are regulated using the RIIO regulatory framework[18], which gives cost and quality incentives over the eight years from 2013 - 2021. The power of the cost incentive is about 0.5 for each of the British transmission owners. This incentive is chosen by OFGEM, the regulator of electricity and gas, based on the transmission owners' proposal of costs. A proposal closer to efficient costs (as determined by the OFGEM) received a higher incentive power (OFGEM, 2016).[19] Table III.1 shows that NGET has an incentive power of 0.4689,

[16] This is done for many distribution networks worldwide, but less for transmission networks, as there are relatively few of them, they depend on many variables (the distribution of generation and load, geographic topography, the attributes and age of the legacy network, population density, etc.), and there is no uniform definition of transmission networks (Joskow, 2014).

[17] In Australia the incentive power is $b = 0.3$ for transmission and distribution system operators (Australian Energy regulator, 2013).

[18] RIIO stands for Revenue = Incentives+Innovation+Outputs

[19] For the period 2005-2006, the system operator was offered a menu of contracts where

while SHE and SPT have 0.5.[20] The same table also shows the results of RIIO in 2015-2016 for all three transmission owners and the system operator. NGET TO spent £644m less than allowed and retained 46.89 % of this as higher allowed revenue.

Table III.1: Results of RIIO in 2015-2016 for the three electricity transmission networks (source: OFGEM (2016)).

m£ 2015-16 Prices		NGET TO	NGET SO	SHE	SPET
Allowed Totex	$\overline{C}$	1,805	137	781	354
Actual Totex	C	1,161	137	524	358
Over-/underspend	$C-\overline{C}$	-644	-1	-257	4
Incentive power	b	0.4689	0.4689	0.5	0.5
Allowed Totex	$R=b\overline{C}+(1-b)C$	1,463	137	652	356

In addition to the above cost incentive, the British electricity networks are also subject to a range of additional incentive in the following categories: safety, reliability, availability, customer satisfaction, connections and environment (OFGEM, 2016). We focus on the reliability incentive.[21] The reliability incentive is expressed in energy not supplied (ENS) [MWh]. Table III.2 shows the ENS of the electricity networks for 2013-2016. The value of lost load (VOLL) is considered to be 16,000 £/MWh in GB, irrespective of consumer groups, time of interruption or duration, as in the CENS regulation of Norway. The target ENS is constant over the eight years of RIIO 2013-2021. Note that ENS is considerably below target, which leads to large rewards to the electricity networks. A possible explanation is that the incentive power of the quality incentive is one. That is, over- or underperformance is not shared with consumers. As predicted by equations (3.20), this leads to too-high quality and effort. From Figure III.3, it can be

a higher cost norm was combined with a lower upside and higher downside sharing factor ($s=1-b$). Option 1: $\overline{C}$ = £480m with $s_{up}=0.6$ and $s_{down}=0.15$; Option 2: $\overline{C}$ = £500m with $s_{up}=0.4$ and $s_{down}=0.2$; Option 3 $\overline{C}$ = £515m with $s_{up}=0.25$ and $s_{down}=0.25$ (Joskow, 2014).

[20] RIIO does not use 'incentive power' b but 'sharing factor' $s=1-b$, defined as $R=\overline{C}+s(C-\overline{C})$.

[21] For an overview of quality of electricity supply in almost all European countries, we refer to (CEER, 2011).

seen that $b = 0.5, b_V = 1$ leads to a quality level that is twice the optimal.

Table III.2: ENS of the GB electricity transmission networks for 2013-2016 (source: OFGEM (2016)).

	NGET	SHE	SPET
Target ENS [MWh]	316	120	225
ENS 2013-2014	135	36	42
ENS 2014-2015	9	106	3
ENS 2015-2016	4	0	14
Cumulative reward [m£]	12.8	3.5	9.9

III.5.3.2 Regulation of gas distribution networks in Britain

The eight British gas distribution networks are also regulated according to the above RIIO regulation (OFGEM, 2017b). Table III.3 shows that the incentive power for the eight British gas distribution networks is between 0.63 and 0.64 over the eight years from 2013 - 2021. This table shows that all distribution utilities underspend and earn a higher allowed revenue.

Table III.3: Results of RIIO in 2015-2016 for the eight gas distribution networks (source: OFGEM (2017b)).

£m 2015-16 Prices		EoE	Lon	NW	WM	NGN	Sc	So	WWU
Allowed Totex	$\overline{C}$	316	303	237	186	261	196	406	254
Actual Totex	C	297	238	226	172	227	165	336	209
Over-/underspend	$C - \overline{C}$	-19	-65	-11	-14	-34	-31	-70	-45
Incentive power	b	0.63	0.63	0.63	0.63	0.64	0.637	0.637	0.632
Allowed Totex	$R = b\overline{C} + (1 - b)C$	309	279	233	181	249	185	380	238

Just as in the RIIO regulatory framework for electricity, gas networks are also subject to a range of additional incentives in the following categories: network safety, customer service, social obligations, network reliability, new connections and environmental protection. Some of them have a pecuniary reward or penalty, others are reputational.

III.5.3.3 Regulation of water and wastewater utilities in England and Wales

Water and wastewater utilities in England and Wales are privately owned and operated. An independent economic regulator (OFWAT) supervises the private utilities and applies various forms of benchmarking or yardstick competition to encourage performance improvements (Decker, 2014). Regulation of quality is of paramount importance as it directly affects human health and consumers are generally not able to assess the safety of the water.

As already explained in section III.5.2, OFWAT uses a Totex-approach for its 2015-2020 regulatory period. This means that all costs of a water utility are regulated under the same revenue cap. OFWAT's 2014 price review for the 2015-2020 period consisted of two steps. First, water utilities had to submit business plans that specified their detailed proposal of objectives ($\overline{C}$ and $\overline{q}$), penalty and reward schemes (V) and efficiency sharing factors ($s = 1 - b$) (OFWAT, 2013). Based on their business plans, water utilities were awarded a 'enhanced' or 'non-enhanced' status. As many water utilities had provided their proposals in words instead of numbers, only two utilities were awarded the enhanced status. In a second step, OFWAT proposed a menu of incentives to the utilities, based on all submitted business plans and historical data. Table III.4 and Table III.5 show respectively the menu choice of the cost incentive for the enhanced and non-enhanced utilities in wholesale water and wastewater services. Water utilities with enhanced status received menus with higher incentive power and higher additional income for equal menu choice. The menu choice indicates how much this choice differs from OFWAT's estimate of efficient costs. For example, a menu choice of 85 means that the company?s expenditure choice is 15% lower than the estimate of efficient costs. An enhanced water utility that chooses the 85 menu receives a cost norm of 96.25% of its allowed expenditure and an additional bonus of 2.55%, while a non-enhanced utility only receives a bonus of 2.3%.

Subsequently, the resulting allowed revenue is complemented with the different quality incentives, such as leakage, compliance with quality standards, unplanned interruptions, Satisfaction with taste and odour, etc. Some of them have a pecuniary reward or penalty, others are reputational.

Table III.4: Menu of cost incentives for enhanced utilities in wholesale water and wastewater services (source: OFWAT (2014)).

Menu choice	80	85	90	95	100	105	110	115
Incentive power b	59%	58%	57%	56%	55%	54%	53%	52
Cost norm $\overline{C}$ [%]	95	96.25	97.5	98.75	100	101.25	102.5	103.75
Additional income [%]	2.55	1.95	1.33	0.68	0	-0.70	-1.43	-2.18

Table III.5: Menu of cost incentives for non-enhanced utilities wholesale water and wastewater services (source: OFWAT (2014)).

Menu choice	80	85	90	95	100	105	110	115	120	125	130
Incentive power b	54%	53%	52%	51%	50%	49%	48%	47%	46%	45%	44%
Cost norm $\overline{C}$ [%]	95	96.25	97.5	98.75	100	101.25	102.5	103.75	105	106.35	107.5
Additional income [%]	2.3	1.76	1.2	0.61	0	-0.64	-1.30	-1.90	-2.70	-3.44	-4.20

III.6 Conclusions

This paper studied a linear cost and quality regulation of monopolies confronted with inelastic demand. This is a suitable assumption for network industries, such as electrical grids, gas pipelines, water supply, (rail)roads and telecommunications. The main regulatory issues in these industries are incentives for quality and cost-reducing effort, instead of quality and quantity. As the literature on quality regulation of a monopolist has so far mainly focused on optimal quantity and quality, this paper complements the literature by studying the monopolist's cost function instead of its demand curve.

The analysis shows that both quality and effort increase with the power of the quality incentive. The effect of the power of the cost incentive is ambiguous but under reasonable assumptions it increases effort and decreases quality. Next, we show for a range of functional forms that under uncertainty, the power of the cost incentive and quality incentive should optimally be equal and below one. Last, three case studies in electricity, gas and water show that hybrid revenue caps with quality incentives are increasingly used in network industries. This paper proposes a simple and straightforward model to study behavior of the monopolist under these regulatory schemes.

By focusing on cost efficiency, we necessarily neglect optimal pricing issues (allocative efficiency) and quantity decisions. Although we have argued that this is a suitable assumption for network industries, it would be interesting to extend the analysis to quantity, quality and effort, probably with specific functional forms to obtain unambiguous results. Another interesting extension of the analysis would be to include multiple separate quality incentives, as is done in reality. The current quality variable can be interpreted as a vector and thus contains many aspects of quality, but this does not allows us to explicitly study trade-offs between quality aspects. These important questions are left for future research.

Acknowledgements

I am grateful for very helpful comments from Fridrik Mar Baldursson, Claude Crampes, Guido Pepermans, Dirk Van Hertem and Frank Verboven. Work is conducted through funding by the GARPUR project of the 7th European Community Framework Programme: http://www.garpur-project.eu/.

III.A Robustness of results with respect to functional forms

Table III.6 shows monopoly quantity and effort depending on functional form assumptions. The first row shows the results of section III.2.2 and III.3.2. Row 2 yields similar results as row 1. For row 3, however, effort decreases with the power of the cost incentive. That is, it is profitable to decrease quality so much that effort can also be decreased. The last column of this table shows that all three functional forms lead to similar results for optimal incentive power under uncertainty. Table III.7 shows the results of comparative statics, which is as prediction by proposition 3.

III.B Backward-looking cost and quality norms

Suppose now that both the regulator and the monopolist have perfect information about current and future parameters, but that the cost norm and

Table III.6: Monopoly quantity and effort depending on functional form assumptions.

	$C(e,q)$	$\psi(e)$	e_M	q_M	Uncertainty: $b^* = b^*_V$
(1)	$\frac{\beta}{e(1-q)}$	$\frac{E}{2}e^2$	$\left(\frac{b\beta b_V V}{E^2}\right)^{0.2}$	$1-\left(\frac{b^2\beta^2 E}{b_V^3 V^3}\right)^{0.2}$	$\frac{1}{1+5(1-\alpha)\bar{\epsilon}}$
(2)	$\frac{(1-e)\beta}{1-q}$	$\frac{E}{1-e}$	$1-\left(\frac{E^2}{b\beta b_V V}\right)^{1/3}$	$1-\left(\frac{b\beta E}{b_V^2 V^2}\right)^{1/3}$	$\frac{1}{1+3(1-\alpha)\bar{\epsilon}}$
(3)	$(1-e)\beta q^3$	$\frac{E}{1-e}$	$1-\frac{27E^2 b\beta}{b_V^3 V^3}$	$\frac{b_V^2 V^2}{9b\beta E}$	$\frac{1}{1+(1-\alpha)\bar{\epsilon}}$

Table III.7: Comparative statics depending on functional form assumptions.

		$\frac{dq_M}{db_V}$	$\frac{dq_M}{db}$	$\frac{de_M}{db_V}$	$\frac{de_M}{db}$
$c''_{eq} < 0$		+	-	-	+
$c''_{eq} = 0$		+	-	0	+
	(1)	+	-	+	+
$c''_{eq} > 0$	(2)	+	-	+	+
	(3)	+	-	+	-

quality norm are determined according to some backward-looking rule (Laffont and Tirole, 1993, p.401). That is, there is a pre-scheduled regulatory process to reset or 'ratchet' the norms based on realized cost and quality of previous years (Joskow, 2008a). Some examples:

$$\begin{aligned}
&\overline{C_t} = C_{t-1} && \text{and} \quad \overline{q_t} = q_{t-1} \\
&\overline{C_t} = \sum_{i=t-N}^{t-1} \frac{C_i}{N} && \text{and} \quad \overline{q_t} = \sum_{i=t-N}^{t-1} \frac{q_i}{N} \\
&\overline{C_t} = kC_{t-1} + (1-k)\overline{C_{t-1}} && \text{and} \quad \overline{q_t} = kq_{t-1} + (1-k)\overline{q_{t-1}}
\end{aligned} \tag{3.33}$$

In that case, a forward-looking monopolist will take into account the effect of its current effort and quality on future norms and thus on its profit. The proof below shows that the monopolist's cost-reducing effort and the quality in its zone are:

$$\begin{aligned}
e_M(\alpha) &= \left(\frac{bV\beta\alpha}{E^2}\right)^{0.2} \\
q_M(\alpha) &= 1 - \left(\frac{b^2E\beta^2}{V^3\alpha}\right)^{0.2}
\end{aligned} \tag{3.34}$$

Where:

$$\alpha = \{1-\delta, 1-\frac{1}{N}(\delta+\delta^2+\dots+\delta^N), \frac{1}{1+\frac{k\delta}{1-\delta}}\} \quad \text{and } 0 \le \alpha \le 1 \tag{3.35}$$

for the norms of equation (3.33). A low α means that the monopolist has a lower incentive for effort and quality. This is because it anticipates the effect of its current actions on future profits, through the backward-looking cost and quality norms. That is, high effort and high quality now will lead to lower cost and higher quality norms in the future. However, this effect can be counteracted by decreasing the incentive power, because its positive effect on the quality level is larger than its negative effect on effort. The optimal incentive power is a trade-off between a higher quality and higher service costs. The result is summarized in the following proposition.

Proposition 5. *If the cost and quality norms are determined according to some backward-looking rule, a forward-looking monopolist will have too low incentives for cost-reducing effort and quality. In that case, the optimal* *incentive power is*

$$b^* = \frac{3\alpha^{0.2}}{2+\alpha^{0.6}} \leq 1 \tag{3.36}$$

where a low α means lower effort and quality incentives.

Proof. We assume that cost-reducing effort has no durable impact on the cost parameter β. The cost function does only depend on a monopolist's current cost-reducing efforts. This is different from Auray et al. (2011), who analyze quality regulation when effort has a long-lasting impact.

For $\overline{C_t} = C_{t-1}$ and $\overline{q_t} = q_{t-1}$, the monopolist's two-period utility is:

$$\begin{aligned} & b\left(\frac{\beta}{e_{t-1}u_{t-1}} - \frac{\beta}{e_t u_t}\right) + V(q_t - \overline{q}_{t-1}) - \frac{E}{2}e_t^2 \\ & + \delta\Big(b\left(\frac{\beta}{e_t u_t} - \frac{\beta}{e_{t+1}u_{t+1}}\right) + V(q_{t+1} - \overline{q}_t) - \frac{E}{2}e_{t+1}^2\Big) \end{aligned} \tag{3.37}$$

Which leads to:

$$\begin{aligned} e_{NO,t} &= \Big(\frac{bV\beta(1-\delta)}{E^2}\Big)^{0.2} \\ q_{NO,t} &= 1 - \Big(\frac{b^2E\beta^2}{V^3(1-\delta)}\Big)^{0.2} \end{aligned} \tag{3.38}$$

The same analysis is valid for $\overline{C_t} = \sum_{i=t-N}^{t-1} \frac{C_i}{N}$ and $\overline{q_t} = \sum_{i=t-N}^{t-1} \frac{q_i}{N}$.

For $\overline{C_t} = kC_{t-1} + (1-k)\overline{C_{t-1}}$ and $\overline{q_t} = kq_{t-1} + (1-k)\overline{q_{t-1}}$, the monopolist's infinite-period utility is:

$$\sum_{t=1}^{\infty} \delta^{t-1}\left(b\Big(\overline{C_t} - \frac{\beta}{e_t u_t}\Big) + V\Big(q_t - \overline{q_t}\Big) - \frac{E}{2}e_t^2\right) \tag{3.39}$$

The monopolist's maximization problem at each period t is:

$$b\Big(\overline{C_t} - \alpha c_t\Big) + V\Big(\alpha q_t - \overline{q_t}\Big) - \frac{E}{2}e_t^2 \tag{3.40}$$

where $\alpha = -1 + \delta k + \delta^2 k(1-k) + \delta^3 k(1-k)^2 + \ldots = \frac{-1}{1+\frac{k\delta}{1-\delta}}$ (Laffont and Tirole, 1993, p.402). As before, this leads to:

$$\begin{aligned} e_{NO,t}(\alpha) &= \Big(\frac{bV\beta\alpha}{E^2}\Big)^{0.2} \\ q_{NO,t}(\alpha) &= 1 - \Big(\frac{b^2E\beta^2}{V^3\alpha}\Big)^{0.2} \end{aligned} \tag{3.41}$$

Including $e_M(\alpha)$ and $u_M(\alpha)$ in equation (3.21) leads to the following total socio-economic cost:

$$\left(E\beta^2V^2\right)^{0.2}\left(\left(\frac{b^2}{\alpha}\right)^{0.2}+\left(\frac{1}{b^3}\right)^{0.2}+\frac{(b^2\alpha^2)^{0.2}}{2}\right) \tag{3.42}$$

The first-order condition of this objective function is

$$\frac{0.4}{\alpha^{0.2}}b^{-0.6}-0.6b^{-1.6}+0.2\alpha^{0.4}b^{-0.6}=0 \tag{3.43}$$

This leads to the following optimal incentive power:

$$b^*=\frac{3\alpha^{0.2}}{2+\alpha^{0.6}}\leq 1 \tag{3.44}$$

□

If $\delta = 0.95$, $N = 5$ and $k = 0.5$, the optimal incentive powers of equation (3.35) are $\{0.76, 0.88, 0.9\}$.

This analysis shows that backward-looking cost and quality norms are suboptimal if the monopolist is forward-looking. A solution is to use benchmark analysis of similar monopolists, as is done for electricity distribution system operators (Estache et al., 2004; Giannakis et al., 2005).

Chapter IV

Cross-Border Exchange and Sharing of Generation Reserve Capacity

IV.1 Introduction

Transmission System Operators (TSOs) are responsible for the security of their transmission system. They use upward and downward reserves to deal with imbalances, caused by unanticipated outages and forecast errors of demand and intermittent supply. Historically, each TSO procured and activated its reserves in its own zone. However – following cooperation in forward markets, the day-ahead market and the intraday market – some TSOs in Europe and the United States recently started cross-border cooperation of reserves. Current cooperation projects are voluntary but the European balancing guideline obliges European TSOs to cooperate on reserves procurement and activation within two to four years after its adoption (ENTSO-E, 2014b). This obligation will increase the number of cross-border balancing projects.

The benefits of cross-border cooperation of reserves have already been studied in the literature. Most of the literature presents case study results. Vandezande et al. (2009) estimate that a Belgium-Netherlands balancing market would have decreased activation costs by 29-44% in 2008, depending on the availability of cross-border capacity. Likewise, Van den Bergh et al. (2017) estimate the benefits of cross-border activation of reserves to be around €25 million a year for a case study of the 2013 Central Western European electricity system (Belgium, France, Germany and the Netherlands). However, they find lower benefits of cooperation if transmission constraints are neglected during cross-border procurement. Farahmand et al. (2012) study the integration of the balancing and procurement markets of Northern Europe, Germany and the Netherlands. They estimate savings of approximately €400 million per year. Gebrekiros et al. (2013) find only a reduction of 2% of procurement costs in a small numerical illustration. van der Weijde and Hobbs (2011) quantify the inter-market benefits using a stylised 4-node network. They find that the benefits of coordinating balancing markets generally exceed unit commitment benefits. In a future with a 45% penetration of renewable generation, Mott MacDonald (2013) estimates operational cost savings of exchange and sharing of reserves on European scale in the order of €3 billion a year. They assume that the increased intermittent and unpredictable generation capacity results in increased volumes of imbalances.

The case study approach in the literature means that there is still a lack of understanding, whether and to what extent TSO cross-border cooperation is economically efficient for each TSO zone and for the region as a whole. The contribution of this paper is to present a general model that analyses three degrees of TSO cooperation in reserves provision. First, we examine autarkic TSO reserve provision - a non-cooperative TSO equilibrium. Next we study the supply efficiency of reserves exchange, where a TSO can acquire reserve capacity in the adjacent TSO area. The last case investigates reserves sharing. Reserves sharing leads to both supply efficiency and dimensioning efficiency. We show that each step in the integration of zones results in progressively lower expected socio-economic costs. We also present a numerical example in order to illustrate the three scenarios. In addition, to get an understanding of their order of magnitude, we estimate the possible efficiency gains of cross-border procurement of generation reserves in Central West Europe and Iberia, based on publicly available procurement data. Lastly, we show that the gains of cooperation are not equally distributed across TSOs. Some TSOs may even experience an increase of procurement costs, which makes voluntary cross-border cooperation harder to achieve. As the European balancing guideline does not specify the details of inter-TSO agreements, there is, however, room for bargaining.

The paper is organised as follows. The next section describes various concepts of electricity balancing, together with types and examples of cross-border balancing mechanisms. Section IV.3 introduces the model and analyses different degrees of cooperation of cross-border reserves procurement. In section IV.4, we estimate the possible efficiency gains of cross-border procurement of generation reserves in Central West Europe and Iberia. Next, section IV.5 studies the implementation of cross-border reserves procurement. Section IV.6 concludes.

IV.2 Electricity balancing

Electricity balancing is the continuous process, in all time horizons, through which TSOs ensure that a sufficient amount of upward and downward reserves are available to deal with real-time imbalances between supply and demand in their electricity transmission system. Imbalances occur due to forecast errors of demand and renewable supply and unforeseen events such as line failures and generation outages. If imbalances between supply and demand persist for a certain period of time, the electricity system could collapse, leading to a blackout.

Most transmission systems consist of different interconnected networks, which are each governed by one TSO. Since system frequency is shared on all voltage levels of a synchronous area, due to the technical characteristics of electricity, power system reliability is considered to be a common good. That is, a non-excludable but rival good. This means that a MW of power can only be used once and that it is technologically difficult to prevent interconnected TSOs from using more than they provide. Underprovision of reserves in one TSO zone could thus lead to a widespread blackout throughout the synchronous area. Therefore, to prevent this 'Tragedy of the Commons', all TSOs in a synchronous area are obliged to provide reserves.

Figure IV.1 shows the two stages of electricity balancing: procurement and activation. First, to ensure that sufficient reserves are available for real-time balancing, TSOs procure an amount of reserves – so-called reserve capacity or balancing capacity – in advance.[1] This reserve requirements, R, is stipulated by network codes and guidelines. To determine the least-cost procurement of reserve capacity to meet the reserve requirement, the TSO holds an open bidding process for each type of reserves[2] for a given future contracting period. Balancing service providers can submit reserve capacity bids, indicating the size [MW] and the price of the bid [€/MW/hour availability]. In the illustration of Figure IV.1, bid 1, bid 2 and part of bid 3 are accepted in the procurement phase to meet a reserve requirement R. Accepted bids are obliged to be available throughout the contracting period. Second, in each activation period[3] of the contracting period the TSO holds another open bidding process where both the procured reserve capacity and available non-procured capacity submit balancing energy bids. Bids are accepted by financial merit order to meet the real-time imbalance or reserve need r_t of the system. Accepted positive bids increase their generation, while accepted negative bids decrease their generation. In return, they receive the activation price p_{act}. In the illustration of Figure IV.1, bid 2, part of bid 3 and an additional non-procured bid are accepted in the activation phase to meet the real-time imbalance r_t.[4]

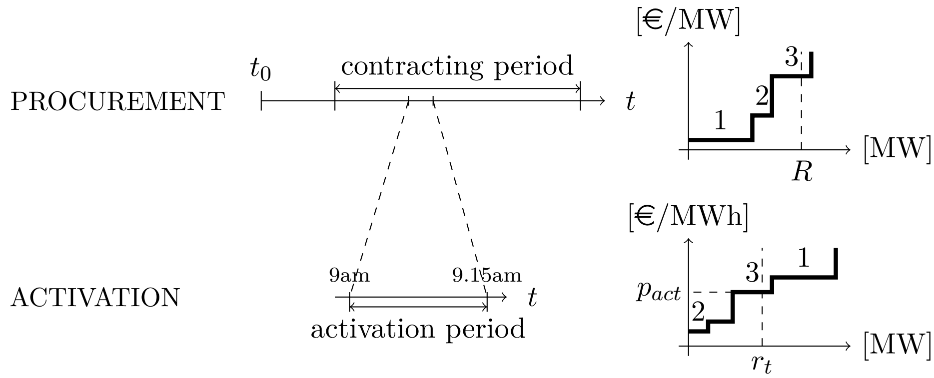
[1] Even network operators with a real-time balancing spot market, like CAISO and Transpower, still procure some reserve capacity in advance. CAISO procures in the day-ahead market and hour-ahead market (Zhou et al., 2016), while Transpower holds a yearly tender for long-term contracts (Transpower, 2013). According to Transpower (2013), the procurement costs are €46.7 million per year.

[2] In Europe, three main categories of reserves exist: (1) Frequency Containment Reserves (FCR), which is used for stabilising the frequency after a disturbance; (2) Automatic and Manual Frequency Restoration Reserves (aFRR and mFRR), which bring the frequency back to its setpoint value; and (3) Reserve Replacement (RR), which replace the active reserves such that they are available to react to new disturbances (ENTSO-E, 2014b). These three types are called primary, secondary and tertiary reserves in North America (Ela et al., 2011).

[3] The activation period, also called settlement period, can be 15 minutes, 30 minutes or 1 hour depending on national market design characteristics. This should be standardized for cooperating TSO zones. According to Neuhoff and Richstein (2016), convergence to the largely used 15 min period is supported by most.

[4] An alternative to merit order activation is pro-rata activation. In that case all procured reserves are activated but in proportion to their relative procurement bid size.

Figure IV.1: Procurement of reserve capacity and activation of balancing energy.



In many TSO zones procurement and activation is more complex than the explanation above. For example, some TSOs co-optimise the market clearing of different types of reserves or assess the reserve capacity bid and the balancing energy bid jointly (50Hertz Transmission GmbH et al., 2014).

Both generation and demand could voluntarily participate in balancing markets, i.e. in both procurement of reserve capacity and activation of balancing energy. However, if the upward reserve need is so large that available reserves are insufficient, the TSO will involuntarily shed load as a last resort to avoid a blackout.

IV.2.1 Cross-border balancing

Under the impulse of increasing renewable energy integration, supranational legislation (ENTSO-E, 2014b), and a general drive for more cost efficiency and reliability, some TSOs have started to coordinate electricity balancing between neighbouring TSO zones. Often cited benefits of cross-border balancing include a more efficient use of electricity generation, including reduced renewable energy curtailment (Mott MacDonald, 2013); reduced reserve needs (NREL, 2011); a higher reliability level (Van den Bergh et al., 2017); internalisation of external effects on neighbouring TSOs (Tangerås, 2012), a standardization of the rules and products, which creates a level-playing field (ENTSO-E, 2014b); and improved market liquidity, which increases competition (Hobbs et al., 2005). In the end, all these benefits decrease the socio-economic cost of balancing. This paper focuses on the first two of the above-mentioned benefits:

(A) **Supply efficiency**: balancing services, both procurement of reserve capacity to meet reserve requirements and activation of balancing energy to meet real-time imbalances, are supplied by the cheapest balancing service providers. That is, if the market is enlarged, expensive balancing services in one part of the market can be substituted for cheaper ones in a different part of the market. The scope for supply efficiency depends on the difference of procurement and activation costs between cooperating TSO zones.

(B) **Dimensioning efficiency**: less procurement of reserve capacity is needed if a TSO in need of capacity can use idle reserve capacity of adjacent TSO zones.

Cross-border cooperation yields benefits both in procurement of reserve capacity and activation of balancing energy. Table IV.1 shows the different degrees of cooperation that are possible in procurement and in activation.

Table IV.1: Degrees of cooperation in cross-border balancing between TSO zones.

PROCUREMENT of reserve capacity	ACTIVATION of balancing energy
To meet the reserve requirements resulting from reserve dimensioning	To meet real-time imbalances resulting from forecast errors and unforeseen events
Autarky: no cross-border cooperation **Exchange**: procure reserves in other zones **Sharing**: multiple zones take into account the same reserves	**Autarky**: no cross-border cooperation **Imbalance netting**: avoid counteracting activation **Exchange**: activate reserves in other zones

First, the three degrees of cooperation in procurement of reserve capacity are autarky, exchange and sharing. Reserves exchange makes it possible to procure part of the required level of reserves in adjacent TSO zones. These reserves are contractually obliged to be available for activation by the contracting TSO and they can only contribute to meeting this TSO's required level of reserves. Reserves exchange changes the geographical distribution of reserves. More reserves are procured in cheap TSO zones and less in expensive TSO zones. Reserves exchange increases supply efficiency by decreasing the procurement costs.

Reserves sharing allows multiple TSOs to take into account the same reserves to meet their reserve requirements resulting from reserve dimensioning.[5] A TSO in need of balancing energy can use this shared capacity, if other TSOs do not. Reserves sharing leads to both supply efficiency and dimensioning efficiency.

Second, the three degrees of cooperation in activation of balancing energy are autarky, imbalance netting and exchange. Imbalance netting avoids counteracting activation of balancing energy in adjacent TSO zones. For example, activating upward reserves in response to a negative imbalance in one TSO zone, and separately activating downward reserves in response to a positive imbalance in another TSO zone, is inefficient since counteracting imbalances naturally net out on synchronous networks. A simple coordination of imbalances could avoid this inefficiency. Imbalance netting is a constrained version of exchange of balancing energy.

Exchange of balancing energy is a further degree of cooperation in activation of balancing capacity. It implies that cooperating TSOs construct a common merit order of balancing energy bids and select the least-cost activation that meets the net imbalance of the joint TSO zone.[6] Imbalance netting and exchange of balancing capacity increase supply efficiency by decreasing the activation costs.

IV.2.2 Examples of cross-border balancing

Balancing and reserve cooperation between TSOs is still in its infancy. However, a few examples of successful cooperation exist in Europe and the United States:

In Europe, ENTSO-E is reviewing a number of pilot projects with the aim to test the feasibility of a multi-TSO cooperation on the cross border procurement of reserve capacity and activation of balancing energy. First,

[5] In practice, reserves exchange and sharing is not limitless. Baldursson et al. (2016b) summarize the limits on reserves exchange and sharing, as imposed by the EU guideline on electricity transmission system operation (European Commission, 2016).

[6] Other market arrangements, like BSP-TSO and an additional voluntary pool, are also possible (Doorman and Van der Veen, 2013).

the International Grid Control Cooperation (IGCC) is a project of imbalance netting of frequency restoration reserves (FRR) to avoid counteracting activation of balancing energy (Just and Weber, 2015). The IGCC was launched in 2012 and currently consists of TSOs from Austria, Belgium, Czech Republic, Denmark, France, Germany, the Netherlands and Switzerland. Second, a part of this group of countries (Austria, Belgium, Germany, the Netherlands and Switzerland) also jointly procure frequency containment reserves (FCR). Third, the Trans-European Replacement Reserves Exchange (TERRE) is established between UK, France, Great Britain, Greece, Italy, Spain, Portugal and Switzerland. The project aims to jointly activate replacement reserves (ENTSO-E, 2016b; Neuhoff and Richstein, 2016). A fourth example of TSO cooperation is the Regulating Power Market (RPM), which was established in 2002 between Denmark, Finland, Norway and Sweden. The RPM is a common merit order of manual frequency restoration reserves (mFRR) activation. Since the Electricity Balancing guideline obligates European TSOs to cooperate on imbalance netting, exchange of balancing services and sharing of reserve capacity, within two to four years after its adoption (ENTSO-E, 2014b, articles 13-20), the number of projects is expected to increase in the future.

In the United States, a cross-border energy imbalance market (EIM) was established between CAISO and PacifiCorp in November 2014. As of 2016 the cross-border EIM consists of five network operators and public utilities in eight states. According to CAISO (2016), the current benefits of the three main participants (CAISO, PacifiCorp and NV Energy) amounted to $88.19 million between 2014Q4 and 2016Q2 and are expected to increase even more in the future with an increased share of renewable generation.

IV.3 Benefits of cross-border reserves procurement

This section studies the benefits of cross-border procurement of reserve capacity. We derive analytical expressions for the optimal level of procured reserves and study the associated cost decreases. Each degree of cross-border cooperation is analysed: autarky, reserves exchange and reserves sharing.

IV.3.1 Model

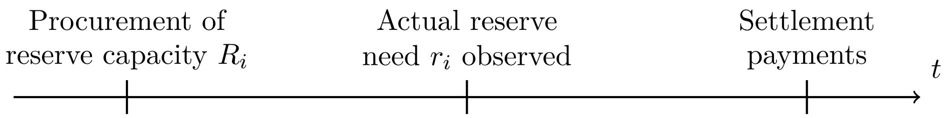
This model studies two TSO zones $i = 1, 2$ that can either not cooperate (autarky), exchange reserves or share reserves. The need for reserves in TSO zone i at a certain instant is denoted by a random variable r_i [MW]. This is the real-time imbalance between supply and demand due to a combination of forecast errors of demand and intermittent supply, and failures of generation capacity or transmission components. We denote the joint probability density function of the reserve needs by $f(r_1, r_2)$ and the marginal density functions of r_1 and r_2 by f_1 and f_2 respectively.[7] The TSO's variable of choice is R_i [MW], the quantity of reserves procured for its own zone i. The contracting period for the procurement of reserve capacity could be e.g. an hour, a week, a month, or a year. In the model we only focus on procurement of upward reserves. Negative reserve procurement is the mirror analysis and its equations are similarly interpreted.

In this paper we are interested in efficiency gains from exchange or sharing of reserve procurement, not efficient activation as such. Hence, the model does not take reserves activation into consideration and we therefore take marginal generation costs to be equal to zero. Costs of procuring R_i of reserve capacity in TSO zone i, however, are not zero and are given by $\gamma_i(R_i)$, with γ_i increasing, smooth and convex.

Figure IV.2 summarizes the order of events. First the TSO at each node i chooses how much reserve capacity R_i to procure. In case of exchange or sharing of reserves, the procurement may entail payments between TSOs. Next, in real time, the actual need for reserves r_i is observed in each node i. The procured reserves will be used to accommodate the reserve needs. In case local reserves are insufficient, TSOs will use exchanged or shared reserves, or, as a last resort, carry out load shedding. Last, settlement payments - if any - are made.

[7] The joint probability density function $f(r_1, r_2)$ will in general depend on the procurement interval and the time to real-time operation.

Figure IV.2: Order of events



IV.3.2 Optimal autarkic TSO reserve provision

We first consider the case where there is no trade or exchange of reserves between zones. Thus, each TSO zone operates as an isolated "island". In reality, network codes and guidelines stipulate the quantity of reserves each TSO zone is required to procure.[8] However, here we pursue an alternative approach by considering the first-best outcome within this setting, i.e. where TSO i procures a quantity of reserves R_i such that expected social surplus in Zone i is maximized.[9] We assume the value of lost load (VOLL - measured in €/MWh) is fixed at v and that electricity demand D_i is price inelastic and also valued at v. Hence, for a given level of reserve needs r_i and procured reserves R_i social surplus is given by consumer surplus net of costs of interruptions (due to unserved demand) and costs of procuring reserves,

$$S_i = vD_i - v\left[r_i - R_i\right]^+ - \gamma_i\left(R_i\right). \tag{4.1}$$

The TSO selects R_i to maximize $E\left[S_i\right]$ with respect to R_i

$$\max_{R_i}\left\{vD_i - v\int_{R_i}^{\infty}\left[r_i - R_i\right] f_i\left(r_i\right) dr_i - \gamma_i\left(R_i\right)\right\} \tag{4.2}$$

Equivalently, since demand is inelastic, the TSO can minimize combined costs of interruptions and reserves, i.e.

$$\min_{R_i}\left\{v\int_{R_i}^{\infty}\left[r_i - R_i\right] f_i\left(r_i\right) dr_i + \gamma_i\left(R_i\right)\right\}. \tag{4.3}$$

[8] Such an exogenous requirement is also standard in reliability management of the day-ahead market, where the N-1 reliability criterion is used instead of balancing the costs of reliability and interruptions (Ovaere and Proost, 2016).

[9] If the reserve requirements of network codes diverge from this first-best optimum (e.g. due to imperfect information or socio-political constraints), costs are higher than in the first-best.

This is the approach we shall use henceforth.[10]

Differentiating (4.3) we derive the following first-order condition for the optimal quantity of reserves R_i^a in autarky:

$$v \Pr\{r_i > R_i^a\} = \gamma_i'(R_i^a). \tag{4.4}$$

The condition (4.4) is very intuitive: reserves should be procured up to the point where the marginal cost of procurement (right-hand side) is equal to the marginal expected cost of interruptions (left-hand side). The higher the variance of the probability density function of reserve needs, i.e. the higher the uncertainty, the more generation reserves need to be procured. However, it is not efficient to procure reserve capacity for every possible level of reserve needs. For those with a very low probability of occurrence, it makes more sense to bear the cost of an interruption than to procure additional reserve capacity. Optimal procurement increases with the VOLL v. The second-order condition for minimum is easily seen to be satisfied.

IV.3.3 Reserves exchange

We now turn to the case of reserves exchange, which as explained earlier, makes it possible to procure part of the required level of reserves in adjacent TSO zones. We assume here that sufficient transmission capacity is available to accommodate the flows arising from use of reserve capacity in adjacent TSO zones and thus neglect any limits transmission capacity constraints would place on reserves exchange (Van den Bergh et al., 2017). That is, there is only load-shedding if $r_i > R_i$, irrespective of where the reserve capacity is procured. To make the setting non-vacuous we assume that procurement costs are not symmetrical so there is a motive for reserves exchange.

This sections shows that exchange of reserves only leads to supply efficiency, not dimensioning efficiency. We study two variants of reserves exchange. First, that the required level of reserves in each TSO zone is the same as in autarky (regulated reserve levels); and second, that it is adjusted

[10] Note that this minimization of procurement costs and expected interruption costs is a classic vendor problem (Arrow et al., 1951), with respectively the cost of purchasing stocks and the penalty for depleted stocks.

in accordance with procurement prices of reserves exchange (locally optimal reserve levels).

IV.3.3.1 Regulated reserves levels

Here we assume, in accordance with the EU guideline on electricity transmission system operation (European Commission, 2016), that the required level of reserves in each TSO zone is the same as in autarky, i.e. R_i^a.

In the first-best solution for this setting the two TSOs jointly minimise total costs of procurement, subject to the constraint on reserves. That is, the cheapest reserve capacity in the two TSO zones is procured first. This amounts to the following constrained cost minimization

$$\min_{R_1,R_2} \{\gamma_1(R_1) + \gamma_2(R_2)\} \text{ s.t. } R_1 + R_2 = R_1^a + R_2^a. \tag{4.5}$$

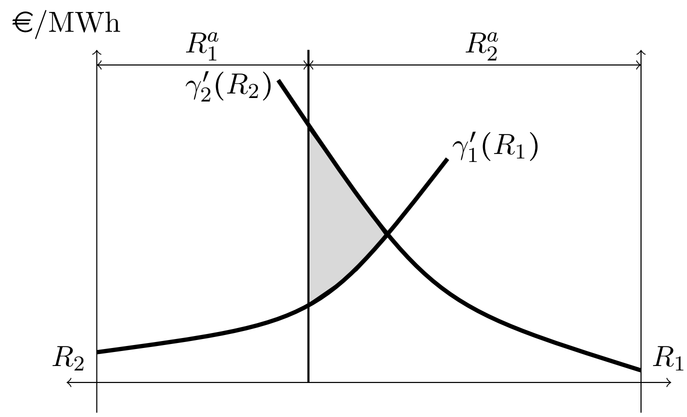
Note that R_i denotes the combined quantity of reserves procured in Zone i by the two TSOs. The side constraint simply says that the overall quantity of reserves procured has to equal the sum of the required reserve levels in the two zones. This minimization problem is easily seen to lead to the following set of equations:

$$\begin{cases} \gamma_1'(R_1) = \gamma_2'(R_2) \\ R_1 + R_2 = R_1^a + R_2^a. \end{cases} \tag{4.6}$$

That is, overall costs are lowest when the marginal cost of reserve procurement is equal in the two TSO zones. Fig. IV.3 shows this cost minimization graphically.

The axis runs from left to right for TSO zone 1 and from right to left for TSO zone 2. The upward sloping lines are the marginal procurement costs in Zone 1 and 2. Clearly, if costs are symmetrical in the two zones, then there is no reason to exchange reserves and the optimal solution is for each TSO to procure reserves within his own zone. If costs are asymmetrical, then there is a rationale for exchange. The grey area in the figure represents the reduction of procurement costs under the optimal procurement of reserves as compared to the costs in autarky where exchange is not possible and each zone supplies its own required reserves.

Figure IV.3: Cost minimization under reserves exchange between two TSO zones



IV.3.3.2 Locally optimal reserves levels

In the previous section we considered reserves exchange where required reserve levels were assumed to be given by regulation.[11] Since the regulatory levels in our model were set so as to match marginal costs of interruptions and reserves, the resulting outcome after opening up for exchange is, however, no longer an optimum: marginal interruption costs no longer match marginal costs of procuring reserves; it will be tempting to lower required reserves in the cheaper zone, where marginal procurement costs have risen, and raise them in the more expensive zone, where they have fallen.

So another scenario is possible when TSOs are allowed to adjust their reserves levels in accordance with prices; this would seem likely to be the tendency over the longer run.

Again, we begin by considering the first-best solution for the present setting. This involves finding the jointly optimal reserve levels, viz. solving

$$\min_{R_1,R_2,R_1^e,R_1^e \text{ s.t. } R_1+R_2=R_1^e+R_2^e} \left\{ \sum_{i=1}^{2} v \int_{R_i^e}^{\infty} \left[r_i - R_i^e\right] f_i\left(r_i\right) dr_i + \sum_{i=1}^{2} \gamma_i\left(R_i\right) \right\} \tag{4.7}$$

where R_j is the amount of reserves procured in Zone j (as before) and R_i^e is the amount of reserves procured by TSO i.

[11]We assumed the regulation to hold TSOs to autarkic levels, even after exchange is allowed, but in principle the regulation could be set at any arbitrary level.

It is readily seen that the optimal solution in this case is determined by the condition that all marginal costs be equal, both across zones and cost types. In other words,

$$\begin{cases} v\,\mathrm{Pr}\left\{r_1 > R_1^e\right\} = v\,\mathrm{Pr}\left\{r_2 > R_2^e\right\} = \gamma_1'\left(R_1\right) = \gamma_2'(R_2) \\ R_1 + R_2 = R_1^e + R_2^e. \end{cases} \tag{4.8}$$

IV.3.4 Reserves sharing

Reserves sharing allows multiple TSOs to draw on the same reserves resources to meet their required level of reserves when it comes to operation. Recall that while exchange of reserves only leads to supply efficiency, reserves sharing leads to both supply efficiency and dimensioning efficiency. As before, we assume that transmission capacity is sufficient to always accommodate the flows arising from use of reserve capacity in adjacent TSO zones. That is, there is only load-shedding if $r_1 + r_2 > R_1 + R_2$.

In our model, reserves sharing amounts to maximizing the surplus of the two zones jointly, in effect uniting them.[12] As before, since we take demand to be inelastic, this is tantamount to minimizing expected costs of interruptions and procurement:

$$\min_{R_1^s, R_2^s} \left\{ v \int_0^{\infty} \int_{R_1^s + R_2^s}^{\infty} \left[r_1 + r_2 - R_1^s - R_2^s\right] f\left(r_1, r_2\right) dr_1 dr_2 - \gamma_1\left(R_1^s\right) - \gamma_2\left(R_2^s\right) \right\} \tag{4.9}$$

The optimal reserve capacities when reserves sharing is allowed, R_1^s and R_2^s, are determined from the following first-order conditions:

$$\begin{cases} v\,\mathrm{Pr}\left\{r_1 + r_2 > R_1^s + R_2^s\right\} = \gamma_1'\left(R_1^s\right) \\ v\,\mathrm{Pr}\left\{r_1 + r_2 > R_1^s + R_2^s\right\} = \gamma_2'\left(R_1^s\right) \end{cases} \tag{4.10}$$

which are derived by differentiation of (4.9) with respect to R_1^s and R_2^s, respectively. The first-order equations imply that marginal costs of reserves procurement are equal to VOLL times the loss of load probability in the two zones together. Clearly, this implies that marginal costs of procurement are

[12] As a simplification, we neglect any limits on reserves sharing, see Baldursson et al. (2016b).

equal at the optimal levels of procurement, $\gamma_1'(R_1^s) = \gamma_2'(R_2^s)$. Hence, the costs of reserves procurement are minimized as in reserves exchange, but for different levels of reserves and, hence, also reliability.

IV.3.5 Efficiency of different degrees of cooperation

To compare the efficiency of the different degrees of cooperation, we need to compute the total costs c^j for each degree of cooperation $j \in \{a, e, l, s\}$. It leads to the following proposition (with a formal proof in Appendix).

Proposition 6. *Each step in the integration of zones results in progressively lower expected socio-economic costs, i.e.* $c^a \geq c^e \geq c^l \geq c^s$.

Proof. Recall that for each degree of cooperation $j \in \{a, e, l, s\}$, R_i^j is the optimal amount of reserves procured in Zone i and c^j is the sum of procurement costs and interruption costs in both TSO zones. By contrast, R_i is the amount of reserves procured by TSO i. Equation (4.11) is the sum of procurement costs and interruption costs with autarky. This minimization determines R_1^a and R_2^a. Adding an additional variable R_1^e leads to equal interruption costs and weakly lower procurement costs in equation (4.12). The inequality is strict if $R_1^e \neq R_1^a$ and $R_2^e \neq R_2^a$. Adding even more variables to allow a trade off between procurement costs and interruption costs causes equation (4.13) to be weakly lower than equation (4.12). Again the inequality is strict if $R_1^e \neq R_1^a$ and $R_2^e \neq R_2^a$. To proof the last inequality, notice that equation (4.13) equals equation (4.14) if the correlation of reserve needs is one. If the correlation is lower than one, both procurement costs and interruption costs decrease.

$$c^a = \min_{R_1^a, R_2^a} \left\{ vE\left[r_1 - R_1^a\right]^+ + vE\left[r_2 - R_2^a\right]^+ + \gamma_1(R_1^a) + \gamma_2(R_2^a) \right\} \tag{4.11}$$

$$\geq c^e = \min_{R_1^e} \left\{ vE\left[r_1 - R_1^a\right]^+ + vE\left[r_2 - R_2^a\right]^+ + \gamma_1(R_1^e) + \gamma_2(R_1^a + R_2^a - R_1^e) \right\} \tag{4.12}$$

$$\geq c^l = \min_{R_1^l, R_1, R_2} \left\{ vE\left[r_1 - R_1\right]^+ + vE\left[r_2 - R_2\right]^+ + \gamma_1(R_1^l) + \gamma_2(R_1 + R_2 - R_1^l) \right\} \tag{4.13}$$

$$\text{with } R_1 + R_2 = R_1^l + R_2^l$$

$$\geq c^s = \min_{R_1^s, R_2^s} \left\{ vE\left[r_1 + r_2 - R_1^s - R_2^s\right]^+ + \gamma_1(R_1^s) + \gamma_2(R_2^s) \right\} \tag{4.14}$$

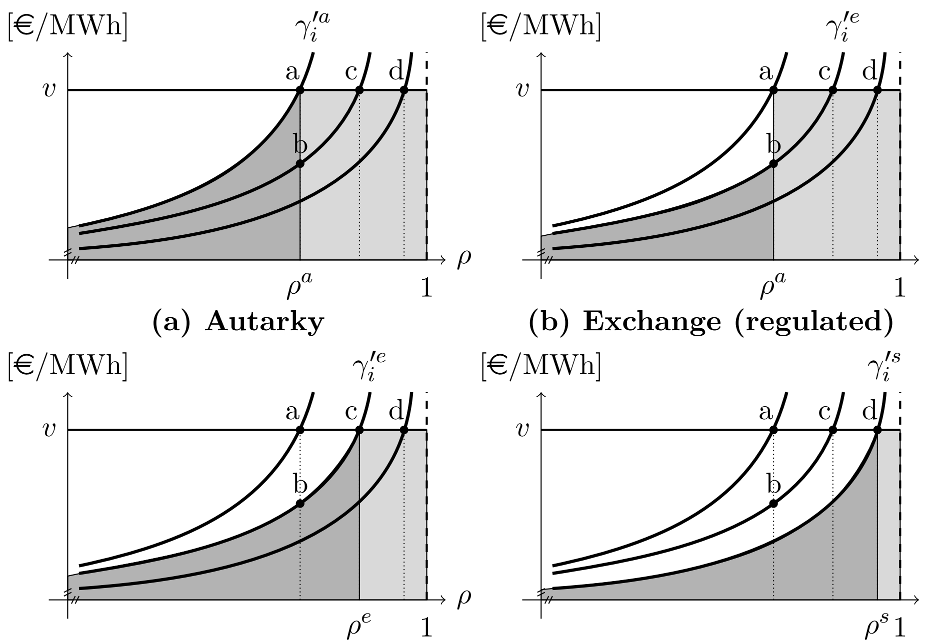
□

This can also be shown graphically. Figure IV.4 shows the socio-economic costs for each degree of cross-border cooperation. The increasing lines are the marginal procurement costs to reach a certain reliability level ρ.[13] As a result, the area below such a line is the total procurement cost to reach reliability level ρ (dark grey area) and interruption costs are $v(1-\rho)$ (light grey area). In aggregate, a higher degree of cooperation leads to lower procurement costs to reach a certain reliability level, i.e. $\gamma^a(\bar{\rho}) > \gamma^e(\bar{\rho}) > \gamma^s(\bar{\rho})$. Figures IV.4a and IV.4b show that moving from autarky to exchange with regulated reserve levels leads to lower procurement costs but leaves interruption costs unchanged, because the reliability level is held fixed. Proceeding to exchange with locally optimal reserve levels (Figure IV.4c) increases procurement costs but less than the decreases of interruption costs. This analysis also shows that moving from autarky to locally optimal exchange has an ambiguous effect on procurement costs because the cost increase of a higher reliability level can exceed the cost decrease of reserves exchange. The cost decrease depends on the cost asymmetry between procurement costs in both TSO zones. Last, reserves sharing

[13] Reliability $\rho \in [0,1]$ can be defined in multiple ways. For example, $\rho = 1 - \text{LOLP}$, where the loss of load probability (LOLP) is the probability of being in a state of the world where some load shedding is needed; or ρ = fraction of total demanded load [MWh] that is supplied to consumers.

(Figure IV.4d) leads to an even higher reliability level and thus interruption costs decrease. As before, its effect on procurement costs is ambiguous and depends on the correlation of reserve needs in TSO zones. The next section presents a numerical illustration of the benefits of cross-border cooperation and studies the comparative statics of the main parameters.

Figure IV.4: Degrees of cooperation: cost efficiency and reliability level.



(a) Autarky **(b) Exchange (regulated)**

(c) Exchange (locally optimal) **(d) Sharing**

IV.3.6 Numerical illustration and comparative statics

The benefits of cross-border exchange and sharing of reserve capacity depend on two parameters: the difference in procurement cost in both TSO zones (c_1 and c_2) and the correlation of reserve needs between TSO zones ($\xi = \text{corr}(r_1, r_2)$). Supply efficiency increases if procurement costs are more asymmetric and dimensioning efficiency increases if reserve needs are less correlated. Figure IV.5 plots the sum of interruption costs and procurement costs with reserves exchange and sharing, relative to the costs in autarky, and shows that the benefits of exchange increase with cost asymmetry (c_1/c_2)

and that the benefits of sharing increase with decreasing reserve need correlation ξ.[14]

Figure IV.5: Relative cost of reserves exchange and reserves sharing, as a function of the cost asymmetry (c_1/c_2) and the reserve needs correlation (ξ).

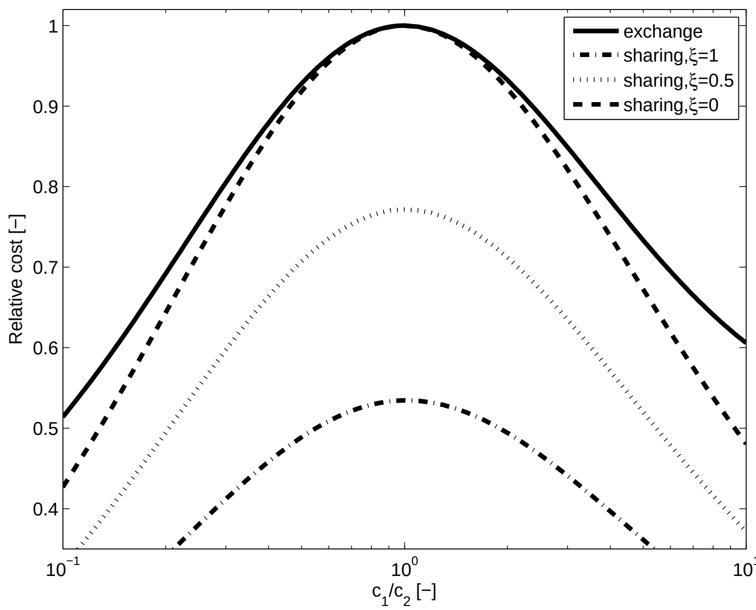


Figure IV.5 illustrates several issues. First, when the two TSO zones have identical procurement costs, no cost arbitrage is possible and exchange of reserve does not yield any cost reduction. However, reserves sharing leads to a lower reserve need and thus a lower cost. Second, when the cost of reserve procurement differs between TSO zones, reserves exchange does yield a cost reduction. For example, when the cost of reserve procurement is higher in TSO zone 1, TSO 1 procures part of its reserve obligation with reserve capacity providers in TSO zone 2. Third, the cost reduction decreases when the reserve needs in the two TSO zones are more correlated. When the reserve needs are fully correlated, reserves sharing yields almost no additional

[14]The probability density functions of reserve needs are jointly normal with correlation ξ, each with a mean of 0 MW and a variance of 100 MW: N(0,100). The cost of reserve procurement in Zone i is $\gamma_i(R_i) = c_i R_i^2$, with $c_1 = c_2 = 1$ at $c_1/c_2 = 1$. The VOLL is 10,000 €/MWh.

cost reduction compared to reserves exchange.

Figure IV.5 also illustrates that the cost reduction increases when reserve procurement costs become more asymmetric and reserve needs are less correlated. With low cost asymmetry and low correlation, reserves sharing yields the major part of the cost reduction, while with high cost asymmetry and a high correlation, reserves exchange yields the major part of the cost reduction. With symmetric costs and high correlation, cross-border cooperation in reserves yields very little cost reduction.

Table IV.2 analyses in detail the reserves, total procurement costs and total interruption costs for the values used in Figure IV.5, with $c_1 = 2$ and $c_2 = 1$. The table shows that sharing reduces the total amount of procured reserves and decreases interruption costs (increases the reliability level). Note that sharing with correlation $\xi = 1$ is equal to exchange with locally optimal reserves.

Table IV.2: Reserves and costs in Zone 1 and 2: $c_1 = 2, c_2 = 1$ (PC = procurement cost, IC = interruption cost).

	R_1	R_2	R_1+R_2	Relative reserves	PC	IC	Total costs	Relative costs
Autarky	23	26	49	100%	1763	479	2242	100%
Exchange	16	33	49	100%	1611	479	2090	93.2%
Sharing $\xi = 1$	17	33	50	100.8%	1638	429	2067	92.2%
Sharing $\xi = 0.5$	15	29	44	88.9%	1273	324	1597	71.2%
Sharing $\xi = 0$	12	24	36	74.4%	891	217	1108	49.4%

In addition to cost asymmetry and the reserve needs correlation, three other parameters influence relative costs of reserves exchange and sharing: VOLL (v), procurement costs, and the relative size of the TSO zones. Table IV.3 compares the relative cost of the base case with a case with higher VOLL, a case with higher procurement costs, and a case where countries differ in size. First, the relative gains of cooperation increase with increasing VOLL, since both the gains of decreased interruption costs and decreased procurement costs are higher. Second, higher procurement costs decrease the relative gains of cooperation. Third, if the TSO zones differ in size[15] the relative gains of cooperation decrease.

Table IV.3: Sensitivity of relative costs [%].

	BASE	$v = 10v_b$	$c_i = 10c_{i,b}$	$\sigma_2 = 6\sigma_b$
Exchange	93.2	92	95.5	96.7
Sharing $\xi = 1$	92.2	91.2	94.0	96.1
Sharing $\xi = 0.5$	71.2	69.9	73.9	85.4
Sharing $\xi = 0$	49.4	48.0	52.4	74.6

IV.4 Estimation of the efficiency gain of cross-border procurement

While the previous section presented a small numerical illustration to show the effect of reserve needs correlation and asymmetry of procurement costs, this section estimates the possible efficiency gain of cross-border procurement of generation reserve capacity between Belgium, France, Germany, the Netherlands, Portugal and Spain. Our estimation differs from earlier studies (see the introduction), because it is not based on simulation but based on actual market data. To our knowledge, the only exception is (Vandezande et al., 2009) who estimate the decreased cost of a Belgium-Netherlands cross-border balancing market in 2008. Our study, however, estimates the decreased cost of cross-border exchange and sharing of generation reserve capacity for 2015-2016 in different subsets of Central West Europe (CWE) and Iberia.

[15] The relationship between the size of a TSO zone and its reserve need standard deviation σ is not linear because larger countries already internalize their imbalance variability. If the correlation of reserve needs between regions of a TSO zone 1 is ξ_1 and this zone is 2^n times larger than an adjacent TSO zone 2, then $\sigma_1 = (\sqrt{2(1+\xi)})^n \sigma_2$. If $\xi_1 = 0.65$, $\sigma_1 = 6\sigma_2$.

IV.4.1 Data

We use price and quantity data of aFRR procurement in Belgium, France, Germany, the Netherlands, Portugal and Spain. These data are publicly available on the ENTSO-E Transparency Platform[16] since the end of 2014. For each considered country i and for each time instant t, these consist of a price p_{it} [€/MWh] and procured capacity R_{it} [MW]. The granularity of time instants goes from hourly (Portugal and Spain) to yearly (France). In Belgium, France and the Netherlands, only the average price of reserve procurement is reported, while Germany, Portugal and Spain report the marginal price of the procurement auction. The price and quantity data are complemented with imbalance data r_{it} [MWh], which has a granularity between 15 minutes and 1 hour. Table IV.4 summarizes the imbalance and procurement data in the considered European countries. The complete dataset consists of 731*6*24 values of r, p and R.

Table IV.4: Summary of available imbalance and procurement data in considered European countries (Source: ENTSO-E Transparency Platform).

	r_t	R_t	Since	Price
Belgium	15’	weekly	01.08.2016	average
		monthly	01.01.2015	average
France	30’	yearly	01.01.2015	average
Germany	15’	weekly	27.06.2011	marginal
Netherlands	15’	monthly	01.01.2016	average
		yearly	01.01.2015	average
Portugal	60’	hourly	13.12.2014	marginal
Spain	60’	hourly	12.12.2014	marginal

We only focus on data for procurement of positive reserves. As prices, procured capacities and imbalances are similar for negative reserves, our estimation of efficiency gains for positive reserves is in the same ballpark as the efficiency gains for negative reserves.

Table IV.5 presents summary statistics of the procurement and imbal-

[16] German data comes from www.regelleistung.net, where marginal data are available.

ance data. For both 2015 and 2016, this table reports the minimum, maximum and average price and quantity. For example, the first row shows that in Belgium the marginal price of procurement[17] is between 17.3 €/MWh and 34 €/MWh, with an average of 23.4 €/MWh. The procured aFRR capacity is between 140 MW and 148 MW, with an average of 141 MW. Germany procures by far the largest amount of aFRR capacity, while Belgium, the Netherlands and Portugal procure the smallest amount of aFFR capacity. Average prices are lowest in Germany and highest in the Netherlands (2015) and Belgium (2016). The last column reports the positive imbalance value r^+ [MW] that is not surpassed in 99.5% of hours. In most countries, regulation requires that procured generation reserve capacity is sufficient to cover the imbalance in 99% of the time.[18] This column shows that Belgium, Germany and the Netherlands procure sufficient capacity in auctions, while France, Portugal and Spain procure less than required. This does not mean that the latter countries procure insufficient capacity in total (e.g. over the counter). Section IV.4.2 explains how we deal we this in our estimation.

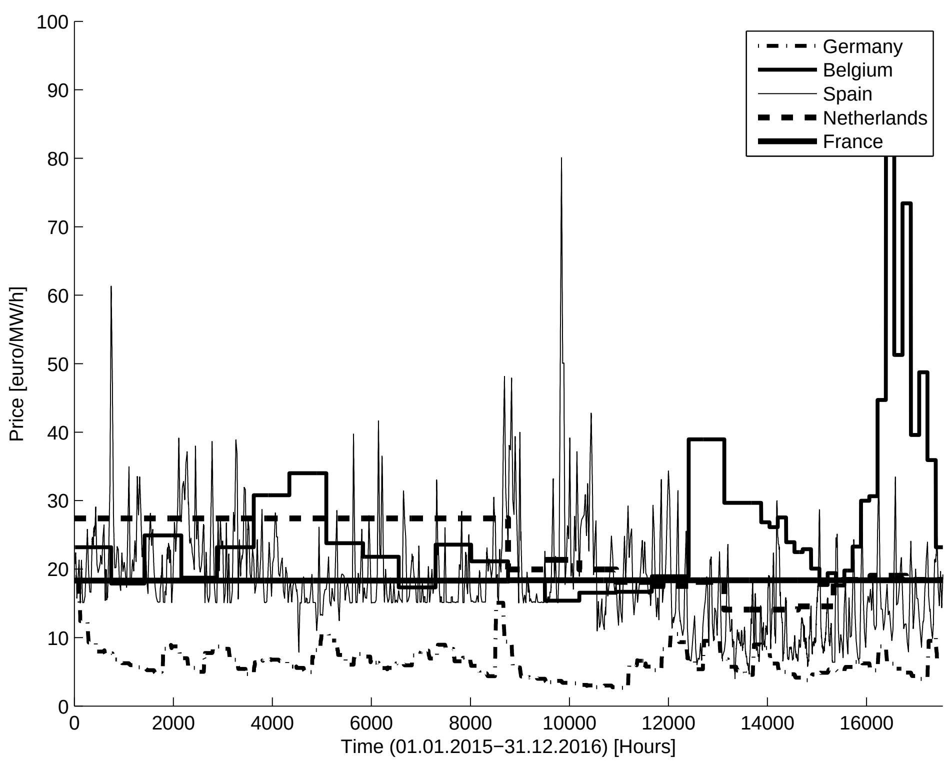
Table IV.5: Summary of aFRR procurement data in considered European countries (Source: ENTSO-E Transparency Platform).

[€/MW/h] and [MW]	Year	p_{min}	p_{max}	p_{av}	R_{min}	R_{max}	R_{av}	$r^+_{99.5\%}$
Belgium	2015	17.3	34	23.4	140	148	141	108
	2016	15.4	87.7	26.8	140	150	142	114
France	2015	18.3	18.3	18.3	500	1177	647	1176
	2016	18.4	18.4	18.4	500	1100	639	1359
Germany	2015	2.58	24.1	7.2	2026	2500	2070	1739
	2016	1.88	24.1	5.6	1973	2500	2014	1541
Netherlands	2015	27.4	27.4	27.4	300	300	300	189
	2016	21.3	14.1	17.8	170	170	170	108
Portugal	2015	5	61.4	20.5	66	322	171	1228
	2016	4	80.1	16.6	56	333	173	1577
Spain	2015	2.1	121	19.6	467	913	685	3846
	2016	0.76	200	15.6	399	927	682	2447

[17] See section IV.4.2 that deals with the calculation methodology.
[18] See (ENTSO-E, 2013, 46(h)).

In addition to the summary statistics of Table IV.5, Figure IV.6 shows the marginal prices of aFRR in Belgium, France, Germany, the Netherlands and Spain for all hours of 01.01.2015 to 31.12.2016. Portuguese prices are not shown because they are close to the prices in Spain.[19] As the hourly data of Germany and Spain is volatile, we report their 24-hour moving average. The prices in France are almost constant throughout the assessed period, while the prices in the Netherlands are constant and above French prices in 2015 but decrease in 2016. This figure also shows that, except for Germany, prices cross constantly. As a result, no single country is the most expensive at all times. In Germany, prices are almost consistently lower than in the other five countries.

Figure IV.6: Marginal price of aFRR in Belgium, France, Germany, the Netherlands and Spain (01.01.2015-31.12.2016)



[19]Prices in Portugal and Spain have correlation coefficient of 0.7 for 2015-2016.

Table IV.6 presents the correlation coefficients between imbalances in the six considered countries. These values are statistically different from zero at the 0.001% level, except for the correlation between Portugal and Germany. However, none of these country-pairs has a high positive correlation, so significant efficiency gains of reserves sharing are possible.

Table IV.6: Correlation coefficients between imbalances in the six considered countries (aFRR).

	France	Netherlands	Germany	Portugal	Spain	Belgium
France	1					
Netherlands	-0.122	1				
Germany	0.016	-0.035	1			
Portugal	-0.016	-0.054	0.005	1		
Spain	0.034	-0.032	0.112	0.040	1	
Belgium	-0.038	0.060	-0.051	0.029	0.018	1

IV.4.2 Methodology

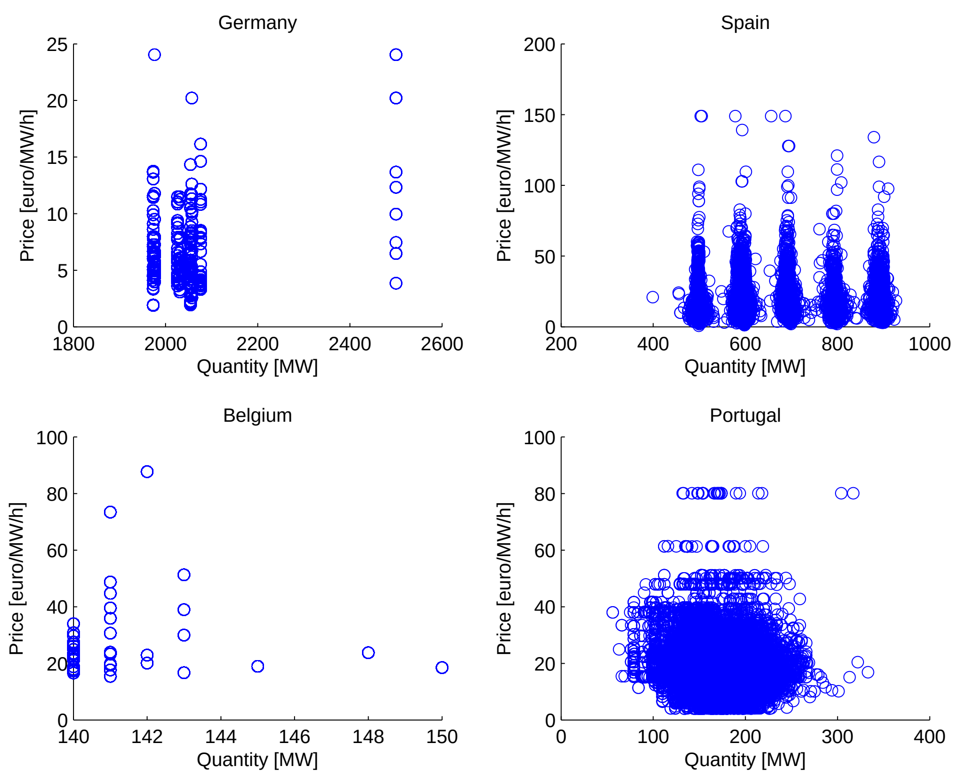
First of all, we need to make an assumption on the functional form of the supply curves of generation reserves. Our only available information is the price-quantity pair for each of the 17544 hours for each country. Figure IV.7 plots these points for Germany, Spain, Belgium and Portugal. These plots clearly show that the supply curve is not constant throughout the period. Therefore, as there is only one price-quantity pair for each hour, we assume that the supply curve is a linear curve between the origin and (R_{it}, p_{it}). That is, for each considered country, the supply curve is different for each hour. For each hour t, the slope of the linear supply curve of country i is:

$$b_{it} = \frac{p_{it}}{R_{it}} \tag{4.15}$$

Secondly, in our dataset some countries report the average price while others the marginal price. As we assume supply to be linear, marginal prices are assumed to be twice the average price.

Thirdly, as before, we do not focus on transmission capacity and assume that it never constraints cross-border cooperation.

Figure IV.7: Scatterplot of procurement price and quantity for Germany, Spain, Belgium and Portugal (01.01.2015-31.12.2016).

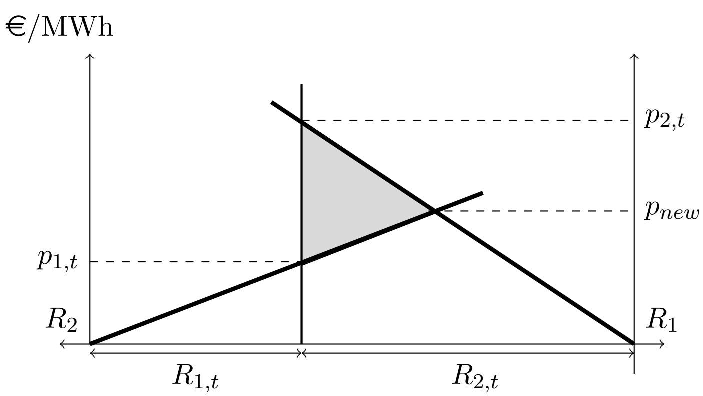


IV.4.2.1 Reserves exchange

The efficiency gains of exchange of reserve capacity can be calculated using equation (4.6) in the case of two countries. Figure IV.8 shows their supply curves and the efficiency gain is represented by the gray area. Generalizing this to exchange of generation reserve capacity between n countries, the common marginal price of procurement p_{new} for each hour t is:

$$p_{new} = \frac{\prod_{i=1}^{n} b_i}{\sum_{i=1}^{n} \prod_{j \neq i}^{n} b_j} \sum_{i=1}^{n} R_i \qquad \textbf{with} \qquad b_i = \frac{p_i}{R_i} \tag{4.16}$$

Figure IV.8: The efficiency gain of reserves exchange between two countries



As the supply slopes are assumed to be linear, the corresponding efficiency gain E from cross-border exchange of generation reserve capacity is:

$$E = 0.5\Big(\sum_{i=1}^{n} R_i p_i - p_{new} \sum_{i=1}^{n} R_i\Big) \tag{4.17}$$

In addition, as ENTSO-E (2014b) imposes that minimally 50% of required aFRR should be in the own country (Baldursson et al., 2016b), we also estimate the above efficiency gains with an additional constraint on allowed exchanged capacity.

We estimate the above equations for each hour separately, which means that the amount of exchanged capacities differs every hour. The total amount of generation, however, is still constant.

IV.4.2.2 Reserves sharing

The gains from sharing of generation reserve capacity between n countries are calculated using the following expression:

$$v \Pr \left\{ \sum_{i=1}^{n} r_i > \sum_{i=1}^{n} R_i^s \right\} = p_{new} \tag{4.18}$$

where p_{new} is calculated from (4.16). The cumulative distribution function of aggregate imbalances in n countries is estimated based on the imbalance data r_{it} of 2015-2016. We see in the data that the probability distribution function of imbalances is a symmetrical bell-shaped curve with mean slightly above zero and fatter tails than the normal distribution.[20]

Again, we estimate the equation for each hour separately, which means that the total procured reserve capacity differs every hour, depending on p_{new}.[21] The higher this price, the lower the procured reserve capacity.

As for reserves exchange, we assume the limits on reserves sharing imposed by (ENTSO-E, 2014b). They state that the amount of procured reserve capacity can not decrease more than 30%, compared to autarky.

The efficiency gain E from cross-border sharing of generation reserve capacity is also calculated from equation (4.17).

As noted before, Table IV.5 showed that the procured aFRR capacity is considerably below imbalances in France, Portugal and Spain, because part of the aFRR capacity is procured outside of the auction. To take this into account, the imbalance data is scaled down such that the margin between $r^*_{99.5\%}$ and R_{av} is the same as for Germany. In this way, we only estimate the possible efficiency gains of the 2015-2016 procurement auctions.

IV.4.3 Results

To make easy comparison possible, the calculations are done separately for 2015 and 2016.

[20] In reality, obviously, it is estimated based on historical data, but since only little imbalance data prior to 2015 is available on the ENTSO-E Transparency Platform, we use 2015-2016 data. This should not greatly influence our estimation results.

[21] To simplify the procurement auction in reality, TSOs might choose $\sum_{i=1}^{n} R_i^s$ for a longer period, which decreases the efficiency gain.

IV.4.3.1 Efficiency gain for reserves exchange

Table IV.7 presents the efficiency gains [million € per year] of reserves exchange between different sets of countries. The second and third column present results for 2015, while the last two columns present results for 2016. This estimation of the efficiency gain is only for procurement of positive aFRR, not activation. Results for the procurement of negative aFRR are in the same ballpark. Also note that we only focus on the decreased cost of procurement. The change of interruption costs (see section IV.3.6) is not assessed. Table IV.7 reports for 2015 efficiency gains of two-country reserves exchange between less than €1 million (Belgium-France and Belgium-Netherlands) and €19 million (France-Germany). Efficiency gains are higher for 2016, except for Germany-Netherlands. Limits on reserves exchange only considerably alter the results for France-Germany and Germany-Netherlands. The efficiency gains increase when more countries are cooperating. Note that the gain of a set of countries is always higher than the sum of all the two-country efficiency gains included in this set. The yearly gains of reserves exchange between Belgium, France, Germany, Netherlands, Portugal and Spain exceed €60 million per year.

Table IV.7: Efficiency gains [M€] from exchange of aFRR for different sets of countries.

[M€]	2015		2016	
	Unconstrained	Constrained	Unconstrained	Constrained
Belgium-France	0.88	0.88	2.89	2.77
Belgium-Netherlands	0.70	0.70	2.73	2.69
Belgium-Germany	7.13	6.36	10.59	9.02
France-Germany	18.94	17.88	24.37	22.32
France-Spain	2.12	2.10	3.48	3.40
Germany-Netherlands	7.17	5.98	2.20	1.89
Portugal-Spain	1.61	1.60	1.88	1.85
France-Portugal-Spain	7.63	7.6	11.19	11.03
Belgium-France-Netherlands	3.58	3.58	3.24	3.11
Belgium-France-Germany-Netherlands	42.40	39.65	39.42	36.04
Belgium-France-Germany-... Netherlands-Portugal-Spain	67.8	64.3	63.06	58.01

IV.4.3.2 Efficiency gain for reserves sharing

Table IV.8 presents the efficiency gains [million € per year] of reserves sharing between different sets of countries. As proven in section IV.3, these are larger than the gains of exchange if the imbalance correlation is below one. Table IV.8 shows that the efficiency gains of sharing are a multiple of these of exchange, which could be explained by the fact that the imbalance correlations are around zero. Table IV.8 reports yearly efficiency gains of two-country reserves sharing between about €10 million (Portugal-Spain) and €75 million (France-Germany). Limits on reserves exchange only considerably alter the results for Belgium-Netherlands, France-Spain and Portugal-Spain. The gains of reserves exchange between Belgium, France, Germany, Netherlands, Portugal and Spain exceed €175 million in 2015 and €150 million in 2016. Note that this efficiency gain is estimated relative to procurement costs in case of optimal reserves procurement in autarky, i.e. according to equation (4.4). As the procured reserve capacity is not necessarily optimal in our data, the change of procurement costs from sharing will be different when compared to current procurement costs.

Table IV.8: Efficiency gains [M€] from sharing of aFRR for different sets of countries.

[M€]	2015		2016	
	Unconstrained	Constrained	Unconstrained	Constrained
Belgium-France	17.23	17.12	21.12	21.10
Belgium-Netherlands	16.63	10.33	13.31	11.37
Belgium-Germany	14.39	14.39	15.05	15.05
France-Germany	60.71	60.71	75.40	75.33
France-Spain	54.61	53.98	61.85	59.37
Germany-Netherlands	24.21	24.21	11.80	11.80
Portugal-Spain	10.58	10.56	9.76	8.96
France-Portugal-Spain	66.12	64.17	71.92	67.52
Belgium-France-Netherlands	43.18	42.22	39.57	39.40
Belgium-France-Germany-Netherlands	98.80	98.46	101.04	100.69
Belgium-France-Germany-... Netherlands-Portugal-Spain	175.70	173.93	150.68	137.93

IV.5 Implementation of cross-border reserves procurement

Whenever TSOs start exchanging and sharing reserves, there are gains and distributional effects. This section first analyses how the benefits of cooperation are distributed. Next, we study what institutions improve the incentive for cooperation.

We consider first the autarkic TSO case. In this case, each TSO can implement a market mechanism to minimize the procurement costs of the reserves required. Next we discuss the distributional effects of reserves exchange via a uniform-price auction. These effects can be negative for one of the parties so that compensation mechanisms need to be put in place to guarantee cooperation. We develop a Nash bargaining game to study the compensation necessary for TSOs to agree an exchange of reserves. This game can be defined for regulated reliability levels as well as for optimized reliability levels.

IV.5.1 Optimal autarkic TSO reserve provision

The analysis of the previous section is based on the premise that the TSO (in lieu of a social planner) has direct control of the available reserves. In a market-based system this is not the case and the reserves have to be procured by some market mechanism. Here we assume a uniform-price auction with the resulting price p_i.[22] The TSO now determines the level of reserves R_i that minimizes the cost of procurement and the cost of interruptions:

$$\min_{R_i} \left\{ v \int_{R_i}^{\infty} [r_i - R_i] \, f_i(r_i) \, dr_i + p_i^a R_i \right\}, \tag{4.19}$$

This results in the first-order condition for the optimal level of reserves.

$$v \Pr\{r_i > R_i\} = p_i^a. \tag{4.20}$$

Generation firms supply the reserves. We assume they do not exercise market power and take prices as given, so generators will bid up to the point

[22] Some TSO zones use pay-as-bid clearing but this is considered to be less preferable (Neuhoff and Richstein, 2016).

where marginal procurement costs equal the reserves price, i.e. where

$$\gamma_i'(R_i) = p_i^a. \tag{4.21}$$

The market equilibrium is determined by (4.20) and (4.21). Clearly, equation (4.4) follows from these two conditions so the market equilibrium coincides with the first-best level of reserves in autarky. In a market implementation the resulting reserves price is p_i^a.

IV.5.2 Reserves exchange

IV.5.2.1 Regulated reserves levels without inter-TSO compensation

Now suppose we are in a more realistic setting where, instead of a joint minimization of costs, each TSO minimizes its own costs, subject to the constraint that regulatory reserve levels must be met. As in the autarkic setting, we assume reserves in each TSO zone are procured by a uniform-price auction and, moreover, that these auctions are run simultaneously. Since exchange is unfettered, prices and marginal procurement costs will be equal in the two zones, i.e.

$$p^e = \gamma_1'(R_1) = \gamma_2'(R_2), \tag{4.22}$$

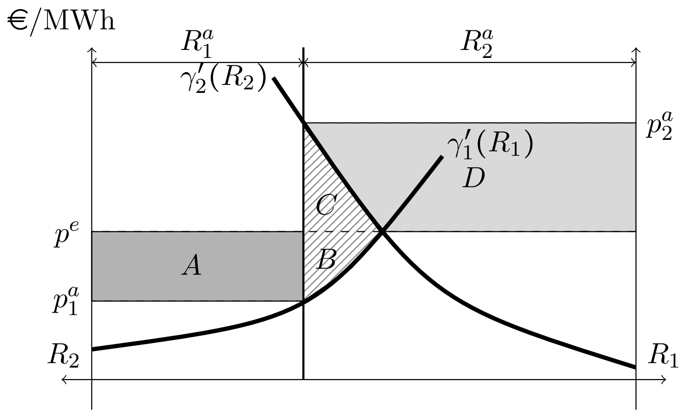
where p^e denotes the price of reserves in exchange, common to the two zones. Comparing (4.22) to (4.6), since each TSO will procure the level of reserves required by regulation, it is clear that the market solution achieves the cost-minimizing outcome.

IV.5.2.2 Inter-TSO transfers to guarantee cooperation

In a transition from autarky to exchange, the reserves price will rise in the cheap zone where marginal procurement costs are lower in autarky than in exchange, and fall in the expensive zone where these costs are higher. Hence, the TSO in the cheap zone will not have an incentive to participate in joint procurement auctions without compensation. Figure IV.9 shows this situation, with Zone 1 being the cheaper and Zone 2 the more expensive. The financial gain of TSO 2 corresponds to area C+D, whereas the loss of

TSO 1 corresponds to area A. TSO 1 can compensate TSO 2 for his loss and retain some surplus provided $C + D > A$.

Figure IV.9: Cost minimization under reserves exchange between two TSO zones. Area A indicates the procurement cost increase of TSO 1; area C+D indicates the procurement cost decrease of TSO 2.



If the cross-border reserves procurement is organized via a uniform-price auction, we need transfers between the TSOs to guarantee cooperation. We will analyse the situation where there are lump-sum transfers.[23] In principle, there are infinitely many solutions to the bargaining game between the two TSOs, as long as a bargaining solution is feasible. Here we use the approach of the Nash bargaining game (Nash, 1953; Binmore et al., 1986) and assume that the autarkic solution is the fallback for both TSOs. Assuming consumers are compensated for interruptions, total costs for TSO i in autarky are $C_i^a = p_i^a R_i^a + L_i$, where $L_i = v \int_{R_i}^{\infty} [r_i - R_i^a] f_i(r_i) dr_i$ are expected interruption costs.[24] We denote the lump-sum side payment from TSO 2 to TSO 1 by x. Similar to (Kolstad, 2005), the side payment can be interpreted as a measure of difficulty to make an agreement.

With exchange the TSOs have the following costs:

$$C_1^e = p^e R_1^a + L_1 + x$$

[23] Another possibility is a distortionary tax on import or export.

[24] Since required reserve levels are the same as in autarky it is in fact irrelevant whether consumers receive compensation. This is no longer the case when reserve levels are allowed to adjust to changed marginal reserve procurement costs.

$$C_2^e = p^e R_2^a + L_2 - x$$

Assuming equal bargaining power of the two TSOs the Nash product is given by

$$N = \left[\left(p_1^a - p^e \right) R_1^a + x \right] \left[\left(p_2^a - p^e \right) R_2^a - x \right] \tag{4.23}$$

and the first-order condition for maximum with respect to x turns out to be

$$x^* = \frac{1}{2} \left[\left(\left(p_2^a - p^e \right) R_2^a - \left(p_1^a - p^e \right) R_1^a \right) \right]. \tag{4.24}$$

That is, the transfer is $0.5(A + C + D)$. The drop in costs for TSO i, going from autarky to exchange with bargaining and side payment is seen to be

$$C_i^a - C_i^e = \frac{1}{2} \left[\left(\left(p_2^a - p^e \right) R_2^a - \left(p^e - p_1^a \right) R_1^a \right) \right] \tag{4.25}$$

The right-hand side of (4.25) is half the net financial surplus resulting from reserves exchange (C + D - A). If one TSO has a stronger bargaining position than the other this result would not be reached. In this case the stronger TSO would gain more of the surplus. The basic result that a positive financial surplus is necessary for a bargaining solution to be feasible would, however, clearly still hold.

The analysis above assumes that a TSO only cares about its procurement costs. In reality, however, a TSO is also concerned about social welfare in its zone. In part because increased costs of reserves procurement are charged to consumers through network tariffs, and therefore do not affect TSO profits. Including this welfare concern into the TSO utility function increases the willingness to cooperate. Suppose that a TSO has a preference $\alpha \in [0, 1]$ for social welfare (SW) and $(1 - \alpha)$ for a decrease of procurement costs (PC). It favours cooperation if:

$$\Delta U_i = \alpha \Delta SW_i - (1 - \alpha) \Delta PC_i \geq 0 \tag{4.26}$$

With a lump sum transfer y the TSOs have the following changes of utility:

$$\Delta U_1 = \alpha \Delta SW_1 + (1 - \alpha)(p_1^a - p^e) R_1^a + y \tag{4.27}$$

$$\Delta U_2 = \alpha \Delta SW_2 + (1 - \alpha)(p_2^a - p^e) R_2^a - y \tag{4.28}$$

where ΔSW_1 equals area B and ΔSW_2 equals area C in Figure IV.9. Assuming equal bargaining power of the two TSOs the Nash product is given by

$$N = [\alpha \Delta SW_1 + (1-\alpha)(p_1^a - p^e)R_1^a + y]\,[\alpha \Delta SW_2 + (1-\alpha)(p_2^a - p^e)R_2^a - y] \tag{4.29}$$

and the first-order condition for maximum with respect to y turns out to be

$$y^* = (1-\alpha)x^* + \alpha \frac{\Delta SW_2 - \Delta SW_1}{2} \tag{4.30}$$

That is, if a TSO also cares about social welfare in its zone, the lump sum transfer is lower, which is an indication that voluntary cooperation is easier (Kolstad, 2005).

Proposition 7. *If a TSO, in addition to procurement costs, also cares about social welfare in its zone, the lump sum transfer needed for cooperation is lower: If* $\alpha > 0$, $y^* < x^*$.

Proof. If $\alpha > 0$, $y^* < x^* \Leftrightarrow \frac{\Delta SW_2 - \Delta SW_1}{2} < x^*$, where $\frac{\Delta SW_2 - \Delta SW_1}{2} = 0.5(C - B)$ and $x^* = 0.5\left[\left((p_2^a - p^e)\,R_2^a - (p_1^a - p^e)\,R_1^a\right)\right] = 0.5(D + C + A)$. Therefore $y^* < x^* \Leftrightarrow A + B + D > 0$. Since areas A, B and D are positive, $y^* < x^*$. □

In Europe, the Electricity Balancing guideline obligates TSOs to cooperate on balancing and reserves procurement. Therefore, TSOs will cooperate even without an incentive to do so. However, in regions without an obligation to cooperate, cost-reducing cross-border cooperation will only materialize if all TSOs reap the benefits of cooperation. This can be ensured with side payments, which can be both the explicit value of our analysis (as in the inter-TSO compensation mechanism) or more implicitly (e.g. distortionary import tariffs or transaction costs to join the cross-border cooperation platform).

IV.5.2.3 Locally optimal reserves levels

In the case of locally optimal reserve levels, not only costs of reserves, but also expected consumer interruption costs will change. Hence, the feasibility of a bargaining solution and side payments will be affected. Basic insights, however, remain the same as in the previous case.

IV.5.3 Reserves sharing

As in the case of reserves exchange there are, in general, distributional consequences of reserves sharing that may make one zone better off and the other worse off, both as regards procurement costs and expected interruptions.[25] Similar to reserves exchange, for incentive compatibility of sharing there will be a minimal side payment from the zone that gains the most to the one that is worse off and a bargaining outcome can be predicted using the Nash bargaining solution. If there is sufficiently low correlation in reserve needs between the two zones, it is, however, possible that the gains from lower interruption costs due to integration outweigh any rise in reserves procurement costs. An extreme example of this is when the two zones have perfectly negatively correlated reserve needs. In this case reserve sharing eliminates any needs for reserve procurement! This is, however, unlikely to be the case in real situations.

IV.6 Conclusions

This paper compares three degrees of TSO cooperation in generation reserves provision: autarky, reserves exchange and reserves sharing. We derive analytically the optimal procurement of reserves in each of the three cases and show that costs decrease with cooperation. The benefits of reserves exchange and reserves sharing depend on cost asymmetry and correlation of real-time imbalance variability between cooperating TSO zones. That is, when TSO zones have highly asymmetric reserve procurement costs but highly correlated reserve needs, reserves exchange already yields a high cost reduction. When TSO zones have fairly equal reserve procurement costs but a low degree of reserve needs correlation, reserves sharing is needed to reap the full benefits of TSO reserves cooperation.

[25] With reserves sharing, assigning procurement costs to TSOs is ambiguous since the decrease depends on the correlation of reserve needs between the TSO zones. In addition, expected interruption costs in each TSO zone depend on how interruptions are shared. For example, if interruptions are shared in equal proportions, the distribution of expected interruption cost is different than if the reserves-providing TSO has priority over the reserves-receiving TSO.

Based on actual 2015-2016 market data of reserves procurement of automatic frequency restoration reserves in Belgium, France, Germany, the Netherlands, Portugal and Spain, we estimate the efficiency gains of exchange and sharing for different subsets of these countries. Cross-border cooperation in these six countries leads to around €60 million per year for exchange and around €150 million per year for sharing.

Our model also shows that cross-border reserves cooperation has distributional effects on TSOs. Some TSOs may even experience an increase of procurement costs, so that voluntary cross-border cooperation requires transfers. We derive the side payments that are needed to induce cooperation and show that cooperation is easier when TSOs care not only about their own costs but also care about social welfare in their zone.

In this paper we focused on the changes of procurement and interruption costs generated by more efficient supply. The true benefits of cross-border cooperation can be higher than presented in our model because of improved market liquidity, internalisation of external effects, and increased market access through standardization of rules and products. In addition, TSOs that are first to cooperate can define the rules and standards of cooperation and have therefore lower transaction and compliance costs.

Acknowledgements

The research leading to these results is partly funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No 608540, project acronym GARPUR. While carrying out this research Fridrik Baldursson has been associated with CREE - Oslo Centre for Research on Environmentally friendly Energy. CREE is supported by the Research Council of Norway.

Chapter V

How Detailed Value of Lost Load Data Impact Power System Reliability Decisions: A Trade-off between Efficiency and Equity

V.1 Introduction

Electricity is the backbone of modern society: we want electricity to be available at all times. However, blackouts and interruptions of electricity consumers occur, because of component outages and uncertainty of demand and intermittent supply. Preventing this requires a more redundant, and thus costly, power system. To keep costs under control, national regulators and transmission system operators (TSOs) aim for an adequate level of reliability (NERC, 2007). That is, a reliability level that balances the costs of reaching a reliability level and the costs of electricity interruptions.

The cost of electricity interruptions is strongly determined by the interruption duration and the value of lost load (VOLL). VOLL is a parameter representing the cost of unserved electricity and is generally expressed in monetary units per kWh or MWh. It is an essential parameter to determine the optimal reliability level of a power system. VOLL is used in many applications such as load curtailment contracts (Joskow and Tirole, 2007), network investment decisions (Electricity Authority, 2013), cost-benefit analyses, quality incentive schemes of transmission and distribution networks[1], energy legislation, and reliability standards[2] (Munasinghe and Gellerson, 1979). Most of these applications simplify the VOLL to a single, constant value.

Precise knowledge of VOLL is paramount to make correct reliability decisions. Various studies have estimated VOLL for different countries and for different interruption characteristics, such as interruption duration, time of interruption, interrupted consumer, location and advance notification. Better-informed reliability decisions are possible by using these detailed VOLL data. Because they provide more information about the benefits of reliability management, they ensure a better balance between the costs and benefits.

The most advanced use of detailed VOLL data to date is the Norwegian cost of energy not supplied (CENS) regulation. In the CENS regulation, TSO and DSO revenue caps depend on the interruption costs in their area. Interruption costs are calculated for different consumer groups, and both the time and duration of interruptions have an effect on interruption costs (Kjolle et al., 2008). The CENS quality regulation is expected to give network operators better incentives to achieve an optimal reliability level. For example, to provide a higher level of reliability to high-VOLL consumers or at high-VOLL moments – e.g. by taking more conservative operating decisions or speeding up restoration times. In the Italian quality regulation of distribution networks, VOLL of residential consumers is set at 10,800 €/MWh, while VOLL of non-residential consumers is set at 21,600 €/MWh (Cambini et al., 2016). Interruptions of non-domestic consumers are thus more costly and therefore network operators have an incentive to provide

[1] In such schemes, a network operator's allowed revenue depends in part on its reliability level. For TSOs, France uses a VOLL of 12,000 €/MWh and Italy a VOLL of 15,000 €/MWh (CEER, 2011).

[2] In Great Britain, a loss of load expectation (LOLE) of 3 hours per year corresponds to a VOLL of 17,000 £/MWh (Newbery and Grubb, 2014).

them a higher level of reliability. However, apart from being used in reliability incentive schemes, available detailed VOLL data are not widely used in reliability decision making.

This paper is the first to assess the impact of using different degrees of VOLL detail in reliability management. We develop a theoretical model that shows the efficiency gains – defined as the (relative) cost decrease – of using a VOLL that differs over time and between consumers. Realizing the full efficiency potential of consumer-differentiated VOLL depends on the technological curtailment possibilities. We make a distinction between perfect curtailment (Crew and Kleindorfer, 1976), random curtailment (Chao, 1983), and spatial curtailment – an intermediate option where a network operator curtails load in regions depending on their VOLL. The theoretical model is illustrated using a numerical example that focuses on expected total system cost of TSOs' operational planning and system operation using different levels of VOLL detail. In addition we study the impact of VOLL differentiation on specific consumer groups, with a focus on equity and social acceptance.

This paper is organized as follows. Section V.2 surveys the growing literature that estimates VOLL as a function of different interruption characteristics for different countries. VOLL data of Norway, Great Britain and the United States are discussed in more detail. Section V.3 studies analytically the efficiency gains of using a VOLL that differs over time and between consumers. Section V.4 expands this analysis to a five-node illustrative network with realistic assumptions on network data, generation plants, intermittent generation, failure probabilities, demand, and demand uncertainty. Section V.5 discusses the results and policy implications. Section V.6 concludes.

V.2 Literature review of detailed VOLL data

VOLL depends on many factors (de Nooij et al., 2009):

- Interruption time: season, day of the week, time of the day;
- Interrupted consumers: residential, commercial, industrial, public;
- Interruption duration;
- Weather at the time of interruption;

- Number of consumers affected;
- Current reliability level;
- Advance notification of the interruption;
- Mitigating measures.

Various empirical studies have estimated VOLL as a function of these different factors. In this section we survey these detailed VOLL studies. We restrict ourselves to studies published since 2007 that estimate the effect on VOLL of at least two interruption characteristics. Table V.1 lists 13 studies and shows the level of VOLL detail for each study.

The table shows that almost all studies estimate VOLL for different consumer types. Some estimate as much as 15 consumer types (Growitsch et al., 2013; Reichl et al., 2013; Linares and Rey, 2013; Zachariadis and Poullikkas, 2012), while others estimate only two or three (Sullivan et al., 2009; Electricity Authority, 2013; London Economics, 2013). Many studies also include the influence of the interruption time on VOLL. Most of them distinguish between time of the day, day of the week and season. In addition, some studies estimate the influence of interruption duration, advance notification and location.

As an illustration, Table V.2 to Table V.4 present detailed VOLL data of Great Britain (London Economics, 2013), Norway (EnergiNorge, 2012), and the United States (Sullivan et al., 2009). These data show VOLL for different consumer groups as a function of season, day of the week, and time of day. The Norwegian data consider four consumer types (residential, industry, commercial, and public) and 36 interruption times (three times of interruption, three days, and four seasons). The British data consider two consumer types and eight interruption times. Finally, the United States' data consider three consumer types and 16 interruption times. All data are expressed in both the home currency and in 2015€/MWh.[3]

[3]Purchasing power parities (OECD, 2016) are used for conversion.

Table V.1: Studies that estimate VOLL as a function of different interruption characteristics.

Country	Consumer type	Time	Duration	Advance notification	Location	Source
Australia	x		x			(CRA International, 2008)
Austria	x	x	x			(Reichl et al., 2013)
Cyprus	x	x				(Zachariadis and Poullikkas, 2012)
Germany	x				x	(Growitsch et al., 2013)
Great Britain	x	x				(London Economics, 2013)
Ireland	x	x			x	(Leahy and Tol, 2011)
Netherlands	x	x			x	(de Nooij et al., 2007)
New Zealand	x	x	x		x	(Electricity Authority, 2013)
Norway	x	x	x	x		(EnergiNorge, 2012)
Portugal	x	x				(Castro et al., 2016)
Spain	x				x	(Linares and Rey, 2013)
Sweden		x	x			(Carlsson and Martinsson, 2008)
United States	x	x	x	x	x	(Sullivan et al., 2009)

All three studies use stated-preference methods to determine the VOLL data.[4] However, comparison of VOLL between countries should be done with care (Mitchell and Carson, 1989) since all stated-preference methods differ to some extent in terms of formulation of questions, cost normalisation factors, scenario designs and data formats and since countries differ culturally.

The British and United States data show VOLL as a single value for each time of interruption. The Norwegian data are displayed differently. Table V.3 shows multipliers for the time of day, day of the week and season. Norwegian VOLL for a particular time is found by multiplying the standard VOLL with the corresponding multipliers:[5]

$$V(c, t(h, d, y)) = V(c) f_h(c, h) f_d(c, d) f_y(c, y) \tag{5.1}$$

$V(c)$ corresponds to the base VOLL per consumer group c, while $f_h(c, h)$, $f_d(c, d)$ and $f_y(c, y)$ are the multipliers to incorporate the effect of respectively the time during the day h (e.g. day vs. night), the type of day d (e.g. week vs. weekend) and the season y.[6]

Comparison of the three datasets shows that residential consumers have a lower VOLL than industrial consumers. On weekdays, VOLL of industrial consumers is between 5 (GB, not winter, not peak weekday) and 300 (US, winter weekday afternoon) times higher than for residential consumers. During weekends, their VOLL is more similar. Residential VOLL in Great Britain is higher and closer to industrial VOLL than in the United States and in Norway. Industrial VOLL is the same order of magnitude in all

[4]Stated-preference methods involve asking consumers their willingness-to-accept (WTA) payment for an outage and willingness-to-pay (WTP) to avoid an outage (contingent valuation or choice experiments), or asking the cost of specific interruptions (direct worth). Several cost estimation methods exist, each of them having its advantages and disadvantages (de Nooij et al., 2007). Best-practice guidelines provide recommendations for correct VOLL estimation (CEER, 2010; Hofmann et al., 2010; Sullivan and Keane, 1995).

[5]This assumes that the effect of time, day and season on VOLL is independent. For example, the relative decrease of VOLL in summer for residential consumers is the same irrespective of the time or day.

[6]The Norwegian data also include the effect of interruption duration on VOLL. In the remainder of this paper we assume VOLL to be linear in duration, while in general VOLL is concave in duration.

three countries, except for small commercial and industrial consumers in the United States, which have a substantially higher VOLL.[7]

The detailed VOLL data of Great Britain, Norway and the United States are further used in the numerical illustration of section V.4, but the level of detail is restricted to consumer type and time of interruption.

Table V.2: Great Britain VOLL as a function of time characteristics and consumer groups (London Economics, 2013, Table 1 and Table 2). (a) is expressed in [2011£/MWh], (b) in [2015€/MWh].

		Not winter				Winter			
		Weekday		Weekend		Weekday		Weekend	
		Peak	Not peak	Peak	Not peak	Peak	Not peak	Peak	Not peak
(a)	Residential	9,550	6,957	9,257	11,145	10,982	9,100	10,289	11,820
	SMEs	37,944	36,887	33,358	34,195	44,149	39,213	35,488	39,863
(b)	Residential	11,093	8,081	10,753	12,946	12,757	10,571	11,952	13,730
	SMEs	44,077	42,849	38,749	39,722	51,284	45,551	41,224	46,306

V.3 Theoretical Analysis

Costs decrease if detailed VOLL data are used instead of one constant VOLL at all times and in all regions. This efficiency gain is shown using a simple model.

Suppose a cost $C(\rho)$ is needed to supply 1 MWh of electricity at reliability level ρ. This reliability cost is constant throughout the year. It is increasing convex in the reliability level and approaches infinity at $\rho = 1$. Reliability $\rho \in [0,1]$ is here defined as:

$$\rho = \frac{\text{total demand - curtailed load}}{\text{total demand}} \tag{5.2}$$

That is, ρ is the fraction of all demanded load [MWh] that is supplied to consumers in a certain period.

[7] Note that VOLL of a consumer type is an average of individual consumers of this type, in between which large differences are possible.

Table V.3: Norwegian VOLL as a function of time characteristics and consumer groups (EnergiNorge, 2012, Table A and Table B).

		Residential	Industry	Commercial	Public
VOLL [2010 NOK/MWh]		5,000	116,000	192,000	170,000
VOLL [2015 €/MWh]		469	10,926	17,984	15,888
Season $f_y(c, y)$	Winter	1	1	1	1
	Spring	0.57	0.87	1	0.67
	Summer	0.44	0.86	1.02	0.51
	Autumn	0.75	0.88	1.06	0.58
Day $f_d(c, d)$	Weekday	1	1	1	1
	Saturday	1.07	0.13	0.45	0.3
	Sunday	1.07	0.14	0.11	0.29
Time $f_h(c, h)$	2 AM	0.4	0.12	0.11	0.43
	8 AM	0.69	1	1	1
	6 PM	1	0.14	0.29	0.31

Table V.4: United States VOLL as a function of time characteristics and consumer groups ((Sullivan et al., 2009, Table 3-10, Table 4-10 and Table 5-11)). (a) is expressed in [2009$/MWh], (b) in [2015€/MWh].

		Summer							
		Weekday				Weekend			
		Morning	Afternoon	Evening	Night	Morning	Afternoon	Evening	Night
(a)	Residential	3,412	2,559	2,428	2,428	4,002	3,018	2,887	2,887
	Small C&I	306,833	372,941	196,500	196,045	188,750	236,621	112,156	110,332
	Large C&I	17,774	24,978	21,054	15,688	12,771	18,191	14,857	11,088
(b)	Residential	2,947	2,210	2,097	2,097	3,457	2,607	2,493	2,493
	Small C&I	265,004	322,100	169,713	169,319	163,019	204,364	96,866	95,291
	Large C&I	15,351	21,573	18,184	13,550	11,030	15,711	12,831	9,576
		Winter							
		Weekday				Weekend			
		Morning	Afternoon	Evening	Night	Morning	Afternoon	Evening	Night
(a)	Residential	2,428	1,706	1,378	1,378	2,821	2,034	1,640	1,640
	Small C&I	423,091	530,688	248,931	244,828	250,299	32,370	135,863	131,760
	Large C&I	14,539	21,360	16,232	12,161	10,035	14,992	10,963	8,231
(b)	Residential	2,097	1,473	1,190	1,190	2,437	1,757	1,417	1,417
	Small C&I	365,415	458,343	214,996	211,452	216,177	279,967	117,342	113,798
	Large C&I	12,557	18,448	14,019	10,503	8,667	12,948	9,468	7,109

The optimal reliability level ρ^* is found by minimizing the sum of reliability costs $C(\rho)$ and interruption costs $(1-\rho)V$:[8]

$$\min_{\rho} \{C(\rho) + (1-\rho)V\} \tag{5.3}$$

This is at the point where marginal reliability costs equal marginal interruption costs:

$$C'(\rho^*) = V \tag{5.4}$$

This first-order-condition shows that VOLL influences the optimal reliability level. Since the reliability cost increases in ρ, a high VOLL calls for a high reliability level and a low VOLL for a low reliability level. For example, if VOLL is higher in winter than in summer ($V_w > V_s$), the reliability level should also be higher in winter than in summer. If a TSO, however, bases its reliability level on the yearly-average VOLL $\bar{V}$, it will aim for a constant reliability level $\bar{\rho}$ throughout the year.[9] As a result, its network is too reliable in summer and not sufficiently reliable in winter. This is shown in Figure V.1, where the reliability levels are found at the intersection of VOLL and marginal reliability cost, which is increasing in ρ. In this figure, the reliability cost is the area below the marginal reliability cost $C'(\rho)$, up to the reliability level ρ, while the interruption cost is the area below the VOLL between ρ and 1.

The sum of reliability costs and interruption costs will be lower if the TSO modifies the reliability level with changing VOLL ($\rho_s < \bar{\rho} < \rho_w$), instead of aiming for a constant reliability level $\bar{\rho}$. This efficiency gain is defined as:

$$[C(\rho) + (1-\rho)V] - [C(\rho^*) + (1-\rho^*)V] \ [€] \tag{5.5}$$

Or

$$1 - \frac{C(\rho^*) + (1-\rho^*)V}{C(\rho) + (1-\rho)V} \ [\%] \tag{5.6}$$

[8] If the reliability cost $C(\rho)$ includes all social costs of reaching a reliability level ρ, the optimal reliability level is also the welfare optimum. If only private TSO costs are included, the optimal TSO value differs from the welfare-optimal reliability level.

[9] Obviously, in reality the reliability cost is not constant throughout the year. For example, if $C(\rho)$ is higher in winter and VOLL is constant, it is optimal to have a lower reliability level in winter than in summer. But for the sake of our argument we restrict our focus here to the change of VOLL over time.

Figure V.1: Efficiency gains if VOLL differs over time.

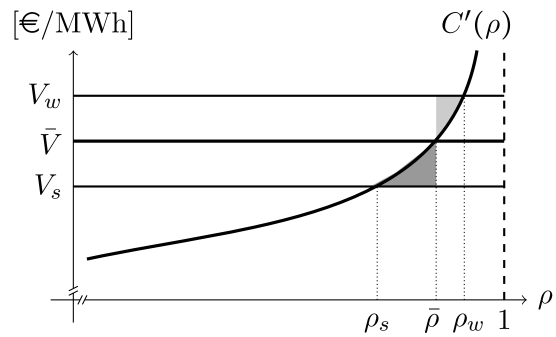
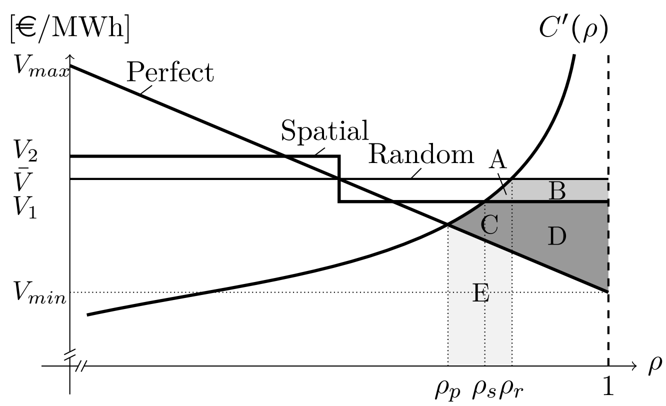


Figure V.1 shows these efficiency gains as the dark grey triangle in summer ($\rho = \bar{\rho}$, $\rho^* = \rho_s$) and the light grey triangle in winter ($\rho = \bar{\rho}$, $\rho^* = \rho_w$). In summer, reliability costs are too high and interruption costs are too low; in winter, reliability costs are too low and interruption costs are too high.

Next, suppose that VOLL is constant throughout the year but differs between consumers. In this case, efficiency gains are achievable by providing low-VOLL consumers with a lower reliability level than high-VOLL consumers. The highest efficiency gain is achieved if demand is curtailed from lowest to highest VOLL (Crew and Kleindorfer, 1976). Perfect curtailment is only possible when the TSO has the technical capabilities to curtail individual consumers. When this is not possible, efficiency gains are still achievable when curtailment is performed first in low-VOLL regions. Spatial curtailment leads to lower interruption costs than random curtailment.

Figure V.2 illustrates the efficiency gains of perfect, spatial, and random curtailment. VOLL is assumed to be uniformly-distributed between V_{min} and V_{max}. This is the downward-sloping line. Moving from random curtailment (with average VOLL $\bar{V}$) to spatial curtailment (with regional VOLLs V_1 and V_2) leads to an efficiency gain equal to the light grey area. This is the sum of lower reliability costs (A) and lower interruption costs (B). The dark grey area is the additional efficiency gain of moving from spatial to perfect curtailment. This is the sum of additional lower reliability costs (C) and additional lower interruption costs (D). Interruption costs are lower because low-VOLL consumers are curtailed first. For spatial curtailment these are consumers in the low-VOLL area 1; for perfect curtailment these are the consumers with the lowest VOLL, in both region 1 and 2. Moving from random curtailment to perfect curtailment, the decrease of reliability costs is thus A+C+E and the net decrease of interruption costs is B+D-E.

Figure V.2: Efficiency gains and reliability level of random, spatial, and perfect curtailment, if VOLL differs between regions.



The regional VOLLs, represented by V_1 and V_2 in Figure V.2, depend on the correlation of VOLL between regions. They differ more if low-VOLL consumers are all concentrated in one region. In that case, the reliability level ρ_s is closer to the optimal reliability level ρ_p and interruption costs of spatial curtailment are lower.

The next section illustrates the theoretical concepts of the current section in a numerical five-node case study.

V.4 Numerical illustration of short-term reliability management

During operation of the electricity system TSOs face many challenges: line outages and generation outages occur, unscheduled loop flows pass through the network, and demand and intermittent supply differ from forecasts. As a result, the TSO takes preventive and corrective actions – such as upward and downward dispatch of generation, phase shifting, transformer tap changing, topological actions and demand curtailment – to ensure that demand

and supply are always balanced without overloading any transmission line. Determining appropriate preventive, corrective and curtailment actions is denoted as short-term reliability management.

V.4.1 Evaluation of short-term reliability management

Short term reliability management consists of two parts: real time operation and operational planning. Both aim at minimizing total expected system costs.

V.4.1.1 Real time operation

When disturbances occur in the power system, the TSO takes corrective actions or curtails load to keep the system in balance. Possible corrective actions a_c^{RT} during real time (RT) operation are generation redispatch, phase shifting transformer tap changing and branch switching. The TSO takes at each time instant t those actions that minimize the cost of corrective actions and the cost of demand curtailment, subject to operational constraints (Van Acker and Van Hertem, 2016).

$$\min_{a_c^{RT}, P_{curt}^{RT}} C_{RT}(v) = \min_{a_c^{RT}, P_{curt}^{RT}} \left[C_{corr}(a_c^{rt}) + P_{curt}^{rt}(c) \cdot v \right] \quad (5.7)$$
$$\text{s.t. operational limits}$$

Load curtailment costs are the product of curtailed load $P_{curt}^{rt}(c)$ and VOLL v. The specification of v depends on the level of VOLL detail:

$$v \in \Upsilon = \{V, V(t), V(n,t), V(c,t)\} \quad (5.8)$$

That is, VOLL is constant (V); VOLL differs over time t ($V(t)$); VOLL is aggregated per node n and differs for all time instants t ($V(n,t)$); or VOLL differs between consumer groups c and over time t ($V(c,t)$). Equation (5.7) shows that different levels of detail in VOLL data change the trade-off between corrective actions and load curtailment and affect which consumers and which regions to curtail. The level of detail has an effect on the choice of corrective actions a_c^{rt} and load curtailment P_{curt}^{rt}, which, in turn, affects total system cost.

V.4.1.2 Operational planning

Real time operation is preceded by the operational planning stage. Operational planning (OP) is executed some time before real-time operation. For example, in day-ahead for the 24 hours of the next day. During operational planning the TSO determines the optimal dispatch of electricity generation, taking into account uncertainties about future real-time states s of the system. The difference between the unconstrained day-ahead market dispatch and the dispatch after operational planning is the cost of preventive redispatch. The TSO determines the dispatch actions a_p that minimizes the sum of preventive redispatch costs $C_{prev}(a_p)$ and expected real-time costs in state s, consisting of the cost of corrective actions $C_{corr}(a_c^s)$ and load curtailment $P_{curt}^s(c) \cdot v$, subject to operational constraints:

$$\begin{aligned} \min_{a_p, a_c^s, P_{curt}^s} C_{OP}(v) = \min \Big[C_{prev}(a_p) + \\ \sum_{s \in S} \pi_s \left(C_{corr}(a_c^s) + P_{curt}^s(c) \cdot v \right) \Big] \\ \text{s.t. operational limits } \forall s \in S \end{aligned} \tag{5.9}$$

where π_s is the probability of occurrence of a possible future real-time state s. The TSO takes into account a set of possible future real-time states S when deciding on its preventive actions a_p. The set S is the Cartesian product of the most probable contingencies and real-time operating states.[10] The latter are conditional upon the forecast values of net demand. As a result, VOLL does not only affect corrective actions and demand curtailment, but also preventive actions of forward-looking TSOs.

Equation (5.3) of our theoretical analysis is a simplified version of equation (5.9). While in the theoretical analysis the TSO chooses the reliability level ρ directly, in our case study it takes a number of preventive (a_p) and corrective (a_c) actions, which lead to a certain reliability level. The reliabil-

[10]Contingencies are considered up to a cumulative probability of 99%, while the possible operating states are approximated by 7 real time realizations of a discrete normal distribution with mean equal to the expected value of total demand at time instant t and a coefficient of variation of 4%.

ity cost $C(\rho)$ of the theoretical analysis includes both the cost of preventive and corrective actions.

V.4.1.3 Evaluation

Performance of short-term reliability management for various levels of VOLL detail is evaluated in terms of expected total cost (ETC). ETC consists of costs of preventive actions, costs of corrective actions and cost of load curtailment.

$$\mathrm{ETC}(v) = \sum_{t \in T} [C_{prev}(a_p(v,t)) + \sum_{rt \in RT} \pi_{rt} \left(C_{corr}(a_c^{rt}(v,t)) \right. \\ \left. + P_{curt}^{rt}(c,v,t) \cdot V(c,t) \right)] \quad \forall t \tag{5.10}$$

Preventive, corrective and curtailment actions

$$[a_p(v,t), a_c^{rt}(v,t), P_{curt}^{rt}(c,v,t)] \tag{5.11}$$

are taken by a TSO based on the available VOLL information, i.e. the level of detail in the VOLL data, $v \in \{V, V(t), V(n,t), V(c,t)\}$. Load curtailment costs are evaluated at the true VOLL of a consumer, $V(c,t)$. ETC is calculated as the expected total cost, averaged over a year.

To calculate the true ETC, the set RT should contain all possible contingencies and all possible real-time operating states. Since this is not feasible in practice, the set RT is bounded, but larger than set S of equation (5.9), which is considered in decision making. The set RT consists of the most probable contingencies up to a cumulative probability of occurrence of 99.6 % and 11 possible real-time realizations of a discrete normal distribution. As the set RT is larger than the set S, reliability management also evaluated in system states that are not considered in advance.

Since more detailed VOLL data lead to better-informed TSO decisions, it is expected that

$$\mathrm{ETC}(V(c,t)) \le \mathrm{ETC}(V(n,t)) \le \mathrm{ETC}(V(t)) \le \mathrm{ETC}(V)$$

In addition to ETC, two other important indicators are the overall reliability level and equity between consumers. The reliability level is expressed in terms of average interruption time (AIT) (Cepin, 2011):

$$AIT = (1-\rho) \cdot 8760 \cdot 60 \qquad [\text{min/year}] \tag{5.12}$$

Equity of the reliability level between consumer groups and consumers at different nodes is evaluated similarly to the Gini coefficient (Atkinson, 1970), but based on the share of total demand that is supplied to the different consumer groups and consumers:

$$G = |1 - (\sum_k (X_k - X_{k-1}) \cdot (Y_k + Y_{k-1})| \tag{5.13}$$

with X the cumulative share of demand, Y the cumulative share of energy not supplied and k an index counting over the groups under comparison, i.e. consumer groups at nodes. The groups are ordered based on decreasing reliability values. A Gini coefficient of 0 means that all consumer groups in all regions have the same reliability level[11]. A Gini coefficient closer to 1 means that all interruptions are concentrated in one or a few consumer groups or nodes. The equity coefficient G indicates how consumers perceive the distribution of reliability between consumer groups in different nodes. It is thus calculated at the aggregated level, not at the level of the individual consumer.

V.4.2 Data

The numerical illustration uses a five-node test system and considers VOLL data of three different countries (Great Britain, Norway and the United States). The same analysis is repeated for each of the countries to determine a range of potential improvements increases in short term power system reliability management if more detailed VOLL data are used. All data are equal for each of the three countries, except the VOLL data.

V.4.2.1 Network

Our illustrative five-node test system is based on the Roy Billinton reliability test system (Billinton et al., 1989), as shown in Figure V.3. Generation is located in node 1 and 2; demand is located in node 2 to 5. Table V.5

[11] Note that the Gini coefficient can not be calculated if reliability is 100% for all consumer groups in all nodes. In that highly exceptional case Gini equals 0 by definition.

shows the reactance (x)[12], capacity and failure probability for the seven transmission lines. All electricity interruptions are assumed to last for 1 hour, implying a linear relationship between VOLL and duration.

Figure V.3: Circuit diagram of the test system.

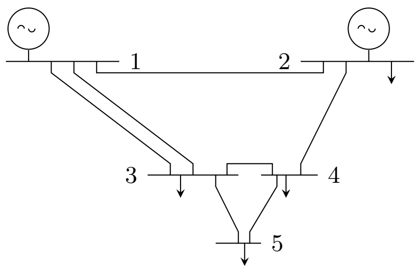


Table V.5: Line data.

From node	To node	x [pu] [MW]	Capacity [MVA]	Failure probab.
1	2	0.48	71	0.0046
1	3	0.18	85	0.0017
1	3	0.18	85	0.0017
2	4	0.6	71	0.0057
3	4	0.12	71	0.0011
3	5	0.12	71	0.0011
4	5	0.12	71	0.0011

V.4.2.2 Generation

The generation park consists of coal-fired power plants with a high marginal cost and wind power plants with a marginal cost near zero, but uncertain availability. Table V.6 summarizes generators' marginal costs and outage probability data. Upward and downward redispatch costs depend on the

[12]The reactance of transmission lines determine the distribution of the power flow in the network: the higher the reactance (compared to other lines), the lower the flow through

Table V.6: Generation data.

Node	Capacity [MW]	Type	C_{marg} [€/MWh]	Failure probab.
1	40	coal	13.83	0.0062
1	40	coal	13.83	0.0062
1	10	coal	13.83	0.0062
1	20	wind	0.04	0.0062
2	40	coal	13.83	0.0062
2	20	coal	13.83	0.0062
2	20	wind	0.01	0.0062
2	20	wind	0.03	0.0062
2	20	wind	0.05	0.0062
2	5	coal	13.83	0.0062
2	5	coal	13.83	0.0062

marginal cost of the generator and differ between the preventive and corrective stage, as shown in equation (5.14). Wind generators are not available for positive redispatch.

$$\begin{aligned} c^+_{prev} &= 1.5 \cdot C_{marg} + 5 \\ c^-_{prev} &= -0.5 \cdot C_{marg} + 5 \\ c^+_{corr} &= 5 \cdot C^+_{prev} \\ c^-_{corr} &= -\frac{1}{5} \cdot C^+_{prev} \end{aligned} \tag{5.14}$$

V.4.2.3 Demand and VOLL

Total system demand is based on the hourly load profile defined for the Roy Billinton Reliability system over a whole year (Billinton et al., 1989). For simplification a year is represented by 6 x 3 x 4 = 72 time instants, each with its probability of occurrence. That is, the set T is the Cartesian product of 6 seasons (early spring, late spring, summer, early autumn, late autumn and winter), 3 days (weekday, Saturday and Sunday), and 4 times

the line.

of day (morning, noon, evening and night). Each temporal case its weighted according to its occurrence. Total system demand at each of the 72 time instants is calculated as the mean over all valid hours. Table V.7 gives the reference share of total demand per node that is attributed to a particular type of customer $LS_{ref}(c)$ together with the share of the total demand at that node. Table V.7 shows that most demand is located in node 3, consisting mostly of residential and commercial demand. Node 4 contains mostly industrial demand, while node 5 contains mostly residential demand.

Table V.7: Demand shares of different consumer groups at different nodes and of demand shares of different nodes in total demand.

	Node	Residential	Industry	Commercial	Public	Total demand share r_T
$LS_{ref}(c,n)$	2	0	0.8	0.2	0	0.125
	3	0.4	0	0.4	0.2	0.5
	4	0.3	0.5	0.1	0.1	0.25
	5	0.8	0.1	0.1	0	0.125

This numerical illustration uses VOLL data from Great Britain (Table V.2), Norway (Table V.3) and the United States (Table V.4). The three datasets consider a different number of consumer types and temporal cases, resulting in different levels of detail. The 72 typical time instants introduced above constitute all temporal cases. In order to unify the data with respect to consumer types, we split consumers into only two categories: residential and non-residential customers. Non-residential customers correspond to the aggregated share of all customers except the residential ones, i.e. large and small C&I combined in the United States and industry, public and commercial combined in Norway. By unifying the test set, we can compare the results in Norway, GB and the US, although their VOLL data have different levels of detail.

The share of residential and non-residential demand in total system demand changes throughout the year. Table V.8 shows the multiplication factors that take this effect into account. The demand share of consumer group c in total system demand at time t is calculated as:

$$LS(c,n,t) = \frac{LS_{ref}(c,n) \cdot f_H(c,h) \cdot f_D(c,d) \cdot f_Y(c,y)}{\sum_{c \in C} LS_{ref}(c,n) \cdot f_H(c,h) \cdot f_D(c,d) \cdot f_Y(c,y)} \tag{5.15}$$

with $c \in$ {residential,non-residential} and t determined by the time of day h, type of day d and time of the year y.

Table V.8: Time dependent multiplication factors for the demand share of different consumer groups.

		Residential	Non-residential
Time $f_H(c,h)$	2 AM	0.7	1.3
	8 AM	1.3	0.7
	2 PM	0.8	1.2
	6 PM	1.3	0.7
Day $f_D(c,d)$	Weekday	0.8	1.2
	Saturday	1.15	0.85
	Sunday	1.3	0.7
Season $f_Y(c,y)$	Winter	1	1
	Spring	0.9	1.1
	Summer	1.1	0.9
	Autumn	1	1

If more detailed VOLL data are used, three cases are distinguished. On the one hand, different consumer groups are considered each with their respective VOLL $v_c = V(c,t)$ and are considered to be curtailable at their respective VOLL. VOLL is on the other hand aggregated per node using a weighted average of the VOLL of the different customer types $v_n = V(n,t) = \sum_{c \in C} LS(c,n,t) \cdot V(c,t)$. In the third case, VOLL is aggregated per time instant using a weighted average of the VOLL at different nodes and the share of total load at that node: $v_t = V(t) = \sum_{n \in N} r_T(n,t) \cdot V(n,t)$.

V.4.3 Results

Our numerical illustration is simulated using a model developed within the GARPUR project [13] (Heylen et al., 2016) and is implemented in AMPL (Fourer et al., 1987) using a MATLAB interface. Probabilistic reliability management is simulated using a probabilistic security constrained DC optimal power flow (Van Acker and Van Hertem, 2016).

Table V.10 shows the relative change of expected total system costs ΔETC for the 5 node test system, which is defined as

$$\Delta ETC = \frac{ETC(v) - ETC(V)}{ETC(V)} \tag{5.16}$$

where v equals VOLL differentiated per consumer group ($v_c = V(c,t)$), VOLL differentiated per node ($v_n = V(n,t)$), or VOLL differentiated per time instant ($v_t = V(t)$), depending on the case under investigation. V represents a constant VOLL for all nodes and consumer groups in all temporal cases.

Before discussing Table V.10, which shows the cost savings in Norway, Great Britain and the United States for different degrees of VOLL differentiation, take a look at Table V.9. This table gives summary statistics of the detailed VOLL data used in our analysis, which will be useful in explaining the results of our analysis. First, the average VOLL (μ) is significantly lower in Norway than in GB and US. Second, when VOLL is constant throughout the country but differing over time (v_t), temporal variation, represented by the coefficient of variation $\frac{\sigma}{\mu}$, is high for Norway, average for US and low for GB. The higher temporal variability in Norway is likely due to the larger relative difference between cold winters and temperate summers. In Norway, the minimum country-wide VOLL is only 255 €/MWh, while it is a hundredfold in both GB and US. The country-wide maximum is between 9,423 and 116,560. This means that optimal reliability will differ substantially over time in Norway, will differ a bit in US and will not change much in GB. Third, when VOLL is allowed to change of time and differentiated between nodes (v_n), the minimum and maximum VOLL will diverge in all three countries. Fourth, when in addition VOLL is differentiated between

[13] `www.garpur-project.eu`

consumers (v_c), minimum and maximum VOLL will diverge even more in all three countries. As a result, the lower the minimum VOLL, the less preventive actions will be taken, as a loss of load is not costly.

Table V.9: Summary statistics of detailed VOLL data in Norway, Great Britain and United States.

		Norway	GB	US
μ		2,095	31,632	57,312
v_t	$\frac{\sigma}{\mu}$	1.1898	0.088	0.4367
	min	255	28,251	27,277
	max	9,423	36,836	116,560
v_n	min	108	15,035	4,832
	max	12,338	51,284	370,364
v_c	min	83	8,081	1,190
	max	19,063	51,284	458,343

With Table V.9 in mind, we can explain the results of Table V.10. First, as expected from the theoretical analysis, cost savings increase with a higher degree of VOLL differentiation. The lower the minimum VOLL that can be curtailed in case of contingencies, the less costly preventive actions are needed. As the minimum and maximum VOLL diverge with a higher degree of differentiation, their cost savings increase accordingly. Secondly, cost savings are large in Norway because its minimum VOLL is close to the cost of preventive and corrective actions. For temporal differentiation (v_t), cost savings are substantial in Norway, low in US, and negligible in GB. The cost savings increase with the level of temporal variation, less with the absolute level of the minimum VOLL, as GB and US have a similar minimum VOLL but different temporal variability. Thirdly, also the cost savings of GB and US increase with more differentiated VOLL data. In that case, it is not the temporal variability but the level of the minimum VOLL that leads to cost savings.

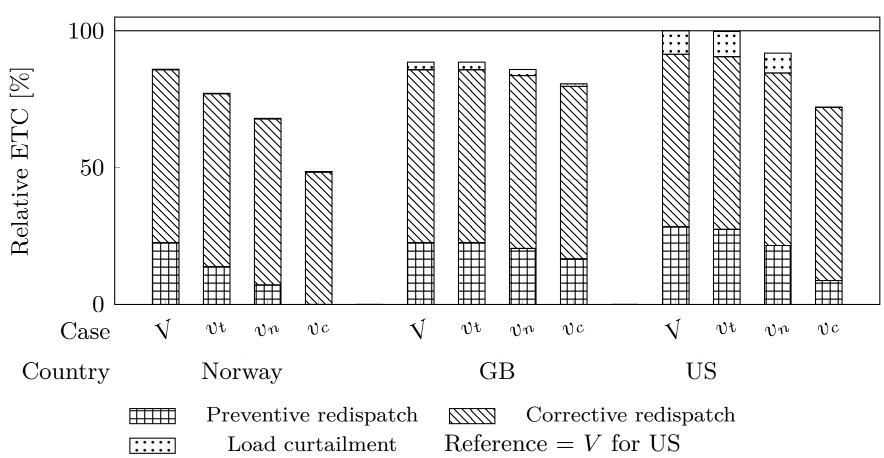
Figure V.4 takes a closer look at how the cost savings of Table V.10 depend on preventive, corrective and curtailment actions. Norway's costs

Table V.10: Relative expected total system cost savings for three countries using VOLL data with different levels of detail.

ΔETC [%]	v_t	v_n	v_c
Norway	-10.68	-20.27	-43.28
GB	-0.01	-3.03	-9.37
US	-0.95	-11.14	-29.52

decrease primarily because it takes less preventive actions, as its cost of curtailing residential consumers is low. GB and US decrease their cost of preventive actions and decrease their curtailment cost when shifting to spatial (v_n) and perfect curtailment (v_c). The largest part of the cost decrease is due to a better trade-off between preventive and corrective actions. Another part is due to better-targeted, and thus lower, curtailment costs, especially for v_c.

Figure V.4: Evolution of cost terms in expected total system cost for different levels of detail of VOLL.



Another important aspect to consider in the discussion is equity of the reliability level between different consumers. If more detailed VOLL data are used and TSOs are able to curtail load based on VOLL, particular consumer groups might experience lower reliability levels. Table V.11

shows the average interruption time per node and consumer group. The last column shows the reliability Gini coefficient, as defined in equation (5.13). Table V.11 shows first that spatial curtailment (v_n) considerably decreases equity. In all three countries, curtailment is almost completely limited to node 5, where low-VOLL residential consumers are located. Second, perfect curtailment (v_c) also decreases equity, but less than spatial curtailment. Curtailment is almost completely limited to residential consumers, as they have the lowest VOLL most of the time. Third, changing VOLL over time (v_t) does not decrease equity. In Norway and US, equity slightly increases; in GB it is constant.

National AIT does not change if more detailed VOLL data are used, except when Norway uses spatial and perfect curtailment based on v_n and v_c respectively. In that case AIT increases because curtailing consumers is cheaper than expensive preventive actions. This is because the absolute level of VOLL is lower in Norway than in GB and US. Because in Norway curtailed (residential) consumers have such a low VOLL, the cost of load curtailment is negligible, as shown in Figure V.4.

V.5 Discussion

The trade-off between efficiency and equity of reliability is an important aspect to consider when introducing more detailed VOLL data. Table V.12 summarizes the reduction of expected total cost (ETC) and the Gini coefficient (G) for the different levels of VOLL detail and for the three countries. If VOLL is equal for all nodes but differs over time, total costs decrease, without a significant effect on equity. In Norway and US equity increases, but this seems to be by chance, as the TSO curtails nodes more randomly[14].

Detailed VOLL data per node v_n or per consumer group v_c, however, have a larger potential for cost savings, but at the expense of increasing inequity. Interestingly, inequity is higher for spatial curtailment than for perfect curtailment. This is because spatial curtailment focuses mostly on the same node (node 5). Perfect curtailment, by contrast, focuses those

[14] Although not completely randomly, because the network topology and the cost of preventive and corrective actions also affect curtailment decisions.

Table V.11: Average interruption time [min/year] (per node and consumer group), consumption weighted average AIT and equity measure (G) for different levels of VOLL detail and different countries.

		Nodes									
	VOLL	2		3		4		5		AIT_{avg}	
Country	Detail	Res	Non res	Res	Non res	Res	Non res	Res	Non res	[min/year]	G
Norway	V	-	1.12	0.31	0.49	1.1	0.37	3.48	16.59	1.91	0.66
Norway	v_t	-	1.04	0.42	0.76	0.66	0.57	3.91	13.59	1.91	0.58
Norway	v_n	-	0.05	0	0	0.16	0.09	23.09	45.54	6.25	0.81
Norway	v_c	-	0.06	14.16	0	127.8	0.03	109.16	0	27.86	0.75
GB	V	-	0.8	0.31	0.31	1.01	0.39	3.5	18.11	1.91	0.7
GB	v_t	-	0.8	0.31	0.31	1.02	0.39	3.5	18.11	1.91	0.7
GB	v_n	-	0	0.05	0.05	0	0	6.52	15.2	1.91	0.82
GB	v_c	-	0.02	1.9	0.01	2.51	0	8.81	0.07	1.91	0.74
US	V	-	1.19	0.92	0.1	0.37	0.72	3.71	15.74	1.91	0.68
US	v_t	-	1.19	0.3	0.49	1.06	0.51	3.94	14.78	1.91	0.64
US	v_n	-	0.11	0.02	0.02	0.02	0.01	4.91	19.95	1.91	0.85
US	v_c	-	0.02	2.45	0	1.87	0	8.48	0.13	1.91	0.73

consumers with the lowest VOLL. Because they are different groups over time, curtailment is more diversified and inequity is lower. This means that if VOLL data is available but perfect curtailment is technologically infeasible, a country should carefully assess if the efficiency gains of spatial curtailment make up for the increased inequity.

Table V.12: Summary table presenting the trade-off between efficiency and equity

	Norway				GB				US			
	V	v_t	v_n	v_c	V	v_t	v_n	v_c	V	v_t	v_n	v_c
ΔETC	0	-10.68	-20.27	-43.28	0	-0.01	-3.03	-9.37	0	-0.95	-11.14	-29.52
G	0.66	0.58	0.81	0.75	0.7	0.7	0.82	0.74	0.68	0.64	0.85	0.73

Increased inequity can also be dealt with by altering network tariffs. If affected consumers have to pay lower tariffs, they would less likely oppose decreased reliability levels. With the introduction of smart meters, reliability levels can even be differentiated per consumer. Reliability contracts can be offered to residential consumers to give them the choice between different reliability levels with corresponding price. Larger consumers can even determine their reliability level on a continuous basis and be paid accordingly. In that case, interruption are not considered as unreliability, as consumers choose the be interrupted in exchange for a payment.

Two issues merit more discussion. First, currently most TSOs do not use even a constant VOLL in their short-term reliability management. Especially not one that is based on extensive VOLL studies. TSOs' reliability decisions are guided by the N-1 criterion. This criterion states that an unexpected outage of a single system component may not result in a loss of load. That is, when a single system component fails, the transmission system should still be able to accommodate all flows without load curtailment. The necessary detailed data (failure rates, forecast errors, wind and solar data, detailed demand and generation data, and, of course, VOLL) are not yet widely available. However, advances in communication and information technologies facilitate gathering this data. With more data available, TSOs can gradually introduce probabilistic methods and interruption costs into reliability management.

Second, actual VOLL strongly depends on the currently perceived reliability level, which is high with currently used reliability management (Munasinghe, 1981). Therefore, VOLL values are in fact not absolute, but conditional upon the perceived reliability level in the country at the moment of the survey. If the reliability level is high, people do not take many actions to prepare for an interruption. While a low reliability level encourages local investments, e.g. in storage or local generation, to prepare for interruptions. If spatial or perfect curtailment is implemented, the reliability level would change for different consumer groups, which in turn changes their VOLL. Due to its low VOLL values, Norway might be mostly impacted by this effect, as people will experience lower reliability levels if exact VOLL data are taken into account in reliability management. Taking into account behavioural feedback effects of VOLL is important, but a lengthy learning process.

V.6 Conclusions

Many empirical studies have estimated how VOLL depends on interruption characteristics – especially consumer type and time of interruption. However, few applications actually use detailed VOLL data to improve power system reliability. A theoretical analysis and a numerical illustration of short-term reliability management both show that incorporating detailed VOLL data leads to considerable efficiency gains. Our numerical illustration leads to potential gains between 3% and 20% when spatial curtailment is used, and between 9% and 43% when perfect curtailment is used[15].

Our analysis showed that this efficiency gain has a downside. Equity of reliability, represented as a Gini coefficient, decreases when more cost effective spatial and perfect curtailment are used. Striking the balance between these opposing objectives is the role of a regulator, based on society's preferences.

When only temporal aspects of VOLL are incorporated, efficiency gains

[15] A back-of-the-envelope calculation, based on 2015 consumption data (ENTSO-E) and the 2015 annual reports of Elia, RTE, Statnett and Terna (only considering operating costs, excluding system losses), leads to an average operating cost of 0.9 €/MWh. Since, total electricity consumption in the ENTSO-E network was 3174 TWh in 2015, this amounts to potential gains between 80 million and 1,200 million per year in the ENTSO-E network.

are lower, but without a significant effect on equity. Therefore, the benefits are clear for countries with much temporal variability of VOLL, like Norway in our numerical illustration.

To reap the benefits of detailed VOLL data in short-term reliability management, two conditions need to be met. First, TSOs need to move away from the currently-used N-1 reliability criterion and move towards probabilistic reliability management. Second, more VOLL studies are needed to improve detailed VOLL data. A widespread roll-out of smart meters have the potential to facilitate the determination of VOLL for different consumer types and different interruption times. Smart meters combined with price-contingent priority rationing contracts will also help to achieve perfect curtailment (Chao and Wilson, 1987; Joskow and Tirole, 2007).

In this paper we focused on the efficiency gains in short-term reliability management. However, considerable gains are also possible in the mid term and long term. A better understanding of interruption costs will lead to better maintenance and system expansion decisions.

Lastly, the increase of intermittent generation will require significant expansions in transmission infrastructure (van der Weijde and Hobbs, 2012). However, the high costs of transmission investments and the difficulties to build new lines in both rural and urban areas could hinder this development (Cohen et al., 2016). This will push power system operation closer to its limits. In such a stressed power system, the use of detailed VOLL data will yield even higher benefits.

Acknowledgements

We are grateful for very helpful comments from Fridrik Mar Baldursson, Guido Pepermans and Frank Verboven. Work is conducted through funding by the GARPUR project of the 7th European Community Framework Programme: http://www.garpur-project.eu/. The work of Evelyn Heylen is funded by the Research Foundation Flanders. All remaining errors are our own.

Chapter VI

An Inequality Indicator of Power System Reliability

VI.1 Introduction

Many decisions of network operators and regulators have an effect on the reliability level of power systems: building new lines, the installation of power flow control equipment, increased penetration of intermittent generation (Heylen et al., 2016a), generation adequacy load-shedding plans (Elia, 2016), the application of new reliability criteria (Heylen et al., 2016b), asset management and maintenance (Bertling et al., 2005), cross-border cooperation on balancing (Baldursson et al., 2016a), etc. However, these decisions do not affect all consumers equally. Some are more affected than others, depending on their location and characteristics. If consumers feel that their reliability level is unfairly low compared to other consumers, they could complain and oppose those decisions that lower their reliability level. Therefore, in addition to measuring the change in costs and the change of the overall reliability level, this paper argues that power system decision makers should also measure the distribution of unreliability among consumers.

To measure the inequality of power system reliability, this paper reformulates the Gini coefficient in terms of energy not supplied. In this way, the adapted Gini coefficient, which is normally used to measure income inequality (Dalton, 1920; Atkinson, 1970; Sen, 1976; Shorrocks, 1980), measures inequality of power system reliability.[1] The proposed inequality indicator summarizes inequality as a value between zero and one. A value of zero means that unreliability is distributed equally among all consumers. The closer the inequality indicator is to one, the more unreliability is limited to a few consumers. We argue that decision makers should not only focus on the impact of their decisions on the aggregate reliability level but also on their impact on the distribution of power system reliability between consumers. A highly unequal or inequitable distribution may lead to public opposition.[2]

The use of the inequality indicator is illustrated in two case studies where the reliability level of consumers is unequally affected. The first case study investigates the inequality of the 2014-2015 load-shedding plan in Belgium. Because of generation adequacy concerns, the Belgian transmission system operator (TSO) ELIA proposed a load-shedding plan that allowed to temporally curtail load in different regions in case of emergency. Public opposition to this plan was strong. People felt that the burden of the load-shedding was placed on a small group of (rural) consumers. The proposed indicator makes it possible to quantify the inequality that results from the load-shedding plan. The second case study illustrates how the inequality indicator can be applied in the performance evaluation of short-term reliability management strategies. Moving towards an alternative reliability management strategy typically involves a trade-off between multiple opposing objectives, of which efficiency and equality are two important ones Ovaere et al. (2016). The inequality indicator allows the comparison of the distribution of reliability among consumers or nodes if TSOs shift from the currently used N-1 reliability criterion to a probabilistic reliability criterion (Karangelos and Wehenkel, 2016b; Capitanescu et al., 2012; Wang and McCalley, 2013).

[1] Throughout this paper we will use the terms 'equality' and 'inequality indicator', but the inequality indicator can also be formulated to represent equity of power system reliability (see section VI.2). Equity is defined as giving everyone what they need or deserve, while equality is defined as treating everyone the same regardless of differences in needs or desert (Konow, 2003).

[2] The inequality indicator based on energy not supplied was firstly applied in a study by Ovaere et al. to analyze how detailed value of lost load data affect reliability decisions taken in probabilistic reliability management and how this impacts the inequality between consumers (Ovaere et al., 2016).

This paper is organized as follows. Section VI.2 formulates the definition of equality, while section VI.3 describes the design of the inequality indicator and discusses its strengths and weaknesses. A first case study in section VI.4 evaluates the inequality resulting from the 2014-2015 load-shedding plan in Belgium, while a second case study in section VI.5 focuses on the comparison of the inequality resulting from short term reliability management based on different reliability criteria. Section VI.6 introduces possible measures to reduce inequality of power system reliability. Finally, section VI.7 concludes.

VI.2 Inequality of power system reliability

First, define a vector **d** that contains the share of demand of each consumer[3] i in the total electrical energy demand of n consumers:

$$d_i = \frac{D_{E,i}}{\sum_{j=1}^{n} D_{E,j}} \tag{6.1}$$

with $D_{E,i}$ the electrical energy demand of consumer i. Next, define a vector **e** that contains the share of energy not supplied (ENS) of each consumer i:

$$e_i = \frac{ENS_i}{\sum_{j=1}^{n} ENS_j} \tag{6.2}$$

with ENS_i the energy not supplied of consumer i. Depending on whether the indicator is used in an ex-ante or ex-post evaluation, resp. expected energy not served (EENS) for a set of events or energy not served (ENS) for a single event or a sequence of events is used.

The following conditions need to be satisfied for vectors **d** and **e**:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} e_i = 100\% \tag{6.3}$$

$$d_e = 0 \Longrightarrow e_i = 0 \tag{6.4}$$

The first condition (6.3) guarantees that all demand and all energy not supplied is distributed over all consumers, while the second condition (6.4) states that consumers without electricity demand cannot have load curtailment.

[3]In this paper, the inequality amongst consumers is used, but similar conclusions can be drawn on substation or regional level.

A distribution of unreliability is considered to be fair or equal if all consumers contribute to the energy not supplied according to their share in total demand:[4]

$$r_i = 1, \forall i = 1..n \text{ with } r_i = \frac{e_i}{d_i} = \text{ inequality ratio} \tag{6.5}$$

If the distribution is not perfectly equal, some consumers i are more ($r_i > 1$) or less affected ($r_i < 1$).

Alternatively, the definition of inequality can be modified by using interruption cost instead of ENS. Interruption costs are the product of a consumer's energy not supplied ENS_i and his value of lost load (VOLL) $VOLL_i$. This formulation states that interruption costs are distributed fairly if the share of each consumer in the interruption cost, either due to the direct consequences of the interruption or due to part of an economical compensation that has to be paid, equals its share in total demand. The inequality ratio in this case equals:

$$q_i = \frac{ENS_i \cdot VOLL_i}{\sum_{j=1}^{n} ENS_j \cdot VOLL_j} \cdot \frac{\sum_{j=1}^{n} D_{E,j}}{D_{E,i}} \tag{6.6}$$

which is equal to 1 in the equality case.

This definition of the inequality indicator is closer to equity of reliability, because VOLL is correlated with need and desert. However, VOLL is not fully correlated with need and desert. For example, poor households may be more in need of reliable electricity supply, but will typically have a lower VOLL than rich households. On the other hand, it makes sense to provide a higher reliability level to hospitals or high VOLL industry. Fairness is a combination of equity and equality, so that both specifications of the inequality indicator are complementary (Konow, 2003). Note that we use the term 'inequality indicator' for both specifications.

[4] In the case of prosumers, both $D_{E,i}$ and ENS_i should be measured from the point of view of the consumer. However, as this is practically not feasible, a formulation from the point of view of the network is a good approximation for low correlation between $D_{E,i}$ and ENS_i. This means that $D_{E,i}$ is defined as the electricity offtake of consumer i, not its net demand, while ENS_i is the energy not supplied by the network.

VI.3 An inequality indicator for power system reliability

Many different inequality indicators have been proposed in the economic literature. These indicators are used to compare income distributions between countries or to verify the impact of certain decisions, such as the introduction of a tax on the distribution of income within a certain country. These indicators have been applied to insurance (Promislow, 1987), education (Vinod et al., 2001), biodiversity (Wittebolle et al., 2009), but not yet to power systems. Based on the definition of inequality provided in the previous section, this section develops an inequality indicator U_{ENS}, which enables the quantification of inequality of power system reliability in a single value.[5]

VI.3.1 Inequality indicators

Various inequality indicators are reported in the literature: the variance, the coefficient of variation, the relative mean deviation (Dalton, 1920), the standard deviation of logarithms, the 20:20 ratio, the Palma ratio, Theil's index (Theil, 1967), the Atkinson index (Atkinson, 1970), the Schutz or Hoover index (Schutz, 1951) and the Gini index (Allison, 1978).

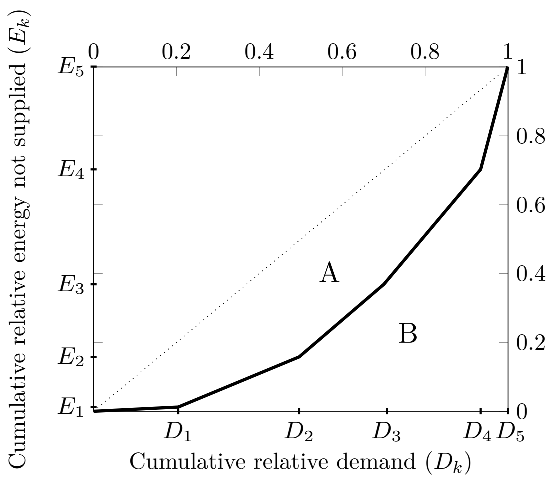
The strengths and weaknesses of each of these indicators have been studied extensively in the economic literature. For example, the variance is not scale invariant[6] and the relative mean deviation fails to satisfy the principle of transfers.[7] In addition, the inequality indicators differ in their sensitivity to transfers: the Palma ratio and the 20:20 ratio particularly focus on the extremes of the distribution, while the Gini indicator focuses on the middle of the distribution (Allison, 1978; Hasenheit, 2014).

[5] The remainder of the paper mainly discusses the inequality indicator in terms of the inequality ratio defined in (6.5). However, the inequality ratio defined in (6.6) can similarly be applied.

[6] Scale invariance ensures that if everyone's ENS or demand is multiplied by a constant value, the degree of inequality remains unchanged (Allison, 1978).

[7] The principle of transfers states that a transfer in the share of ENS s from a consumer i to a consumer j should decrease the value of the inequality indicator if $r_i > r_j$ and $\frac{e_i - s}{d_i} \geq \frac{e_j + s}{d_j}$ (Dalton, 1920).

Figure VI.1: Lorenz curve in terms of energy not served. The line of equality is dotted.



Since the perfect inequality indicator does not exist, we propose to use the Gini indicator, because it is the most widely used. One of the reasons for its popularity is that it is easy to understand how to compute the Gini index based on Lorenz curves.

VI.3.2 Lorenz curves

The distribution of reliability between consumers can be represented in a Lorenz curve. A Lorenz curve plots the cumulative share of demand D_k with respect to the cumulative share of energy not supplied E_k (Fig. VI.1), with all consumers ranked according to an increasing inequality ratio. The inequality ratio represents the slope of the different pieces of the piecewise linear Lorenz curve.

If the distribution of reliability is completely fair (i.e. when $r_i = 1 \ \forall i = 1..n$), the Lorenz curve is a straight line with coefficient of direction equal to 1, as illustrated by the dotted line in Fig. VI.1. If the distribution of reliability is not completely fair, the Lorenz curve will be below the equality line, as illustrated by the bold line in Fig. VI.1. The closer the Lorenz curve is to the equality line, the more equal the distribution of reliability.

VI.3.3 The proposed Gini-based inequality indicator

The proposed Gini-based inequality indicator U_{ENS} of power system reliability is defined as the ratio of the surface area between the line of equality and the Lorenz curve (A) over the total surface area under the line of equality (A+B):

$$U_{ENS} = \frac{A}{A+B} \tag{6.7}$$

Surface area B can be calculated using the surface areas of the trapezoids under each of the pieces of the piecewise linear Lorenz curve. This leads to the following formula for U_{ENS}:

$$U_{ENS} = 1 - \sum_{k=1}^{n} (D_k - D_{k-1})(E_k + E_{k-1}) \tag{6.8}$$

with D_k the cumulative proportion of relative demand ($D_k = \sum_{i=1}^{k} d_i \ \forall k = 1..n$, $D_0 = 0$ and $D_n = 1$) and E_k the cumulative proportion of relative ENS ($E_k = \sum_{i=1}^{k} e_i \ \forall k = 1..n$, $E_0 = 0$ and $E_n = 1$). The consumers i are ranked such that $r_i \leq r_{i+1}$.

VI.3.4 Caveats of the proposed inequality indicator

The main strength of an inequality indicator is that the extent of inequality is summarized as a single value between zero and one. This allows for a simple assessment of the perceived fairness of particular power system decisions. However, this is also the main weakness of the inequality indicator. Aggregating the distribution of reliability into a single value reduces the informational content. For example, the indicator does not capture where the inequality actually occurs in the distribution.[8] Two very different distributions of unreliability can have the same indicator value. Moreover, it is difficult to attribute a practical meaning to a particular value of the indicator, but it is useful in comparison with a well-known reference case or to compare the performance of different power system decisions.

[8] Although the position of each consumer with respect to the equality situation cannot be directly derived from the inequality indicator, it can be obtained based on the inequality ratios r_i calculated per consumer i.

VI.4 Case study I: Controlled load-shedding plans

TSOs with insufficient generation or transmission capacity have the capability and authority to carry out controlled load shedding to prevent uncontrolled failures and blackouts. For example, NERC requires American balancing authorities and transmission operators to have automatic under-frequency load shedding plans (NERC, 2011a) and manual load shedding plans (NERC, 2011b), while ENTSO-E requires European TSOs to have automatic under-frequency control schemes (ENTSO-E, 2016a). TSOs are free to choose which loads to shed in case of emergency, except for high priority significant grid users who should never be shed. However, TSOs generally choose a subset of consumers, which creates public concerns, because people might feel unequally treated.

This case study examines the load-shedding plan that was proposed by the Belgian TSO for the winter 2014-2015.[9] Public opposition to this plan was large, because people felt that the burden of the load-shedding fell on a subset of consumers, while at the same time the benefits accrued to all consumers. To quantify the perceived fairness, this section calculates the inequality indicator of the load-shedding plan.

VI.4.1 Data and assumptions

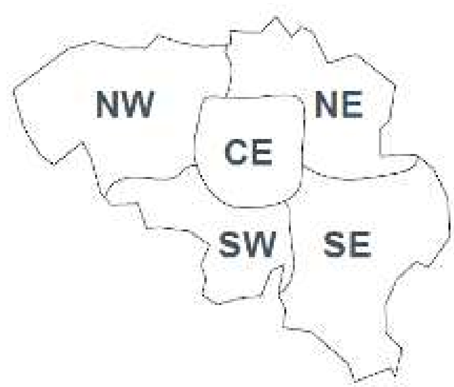
The Belgian load-shedding plan for the winter of 2014-2015 divided Belgium in 5 zones and each zone was further divided into 6 slices.[10] Each slice corresponded to 520 MW of sheddable power, resulting in a total foreseen sheddable load of 3120 MW, as summarized in Fig. VI.2. During load shedding, one of these slices of 520 MW is disconnected according to a rotation system. Slices within a particular zone are determined based on their geographical location, in order to guarantee geographical spreading, and on their value of lost load (VOLL), as rural areas with lower population density and less critical electrical equipment are preferred above urban areas.

[9] At that time, Belgian system adequacy was low due to retirement and mothballing of conventional power plants, supplemented by the unforeseen closure of three large nuclear units as a result of indications of micro-cracks in the reactor vessels.

[10] A recent update of the load-shedding plan uses 8 slices each corresponding to 500 MW up to 750 MW instead of 6 slices of 520 MW Elia (2016).

Figure VI.2: The division of Belgium into zones and slices for the load shedding plan in the winter of 2014-2015 (Data: ELIA).

Slices	NW	NE	CE	SW	SE	TOTAL
1	130	130	130	65	65	520 MW
2	130	130	130	65	65	520 MW
3	130	130	130	65	65	520 MW
4	130	130	130	65	65	520 MW
5	130	130	130	65	65	520 MW
6	130	130	130	65	65	520 MW
7'	8000 MW					
7"	2000 MW					
TOTAL						13120 MW



Total system load is considered to be 13120 MW, which means that 10 GW of load is never affected. These consumers are assumed to be in slices 7' and 7". Slice 7' represents densely populated areas with 8 GW of low-VOLL consumers, while slice 7" represents 2 GW of critical high-VOLL consumers.[11]

VI.4.2 Results

Table VI.1 gives the inequality indicator U_{ENS} between the slices of the controlled load-shedding plan after 1 up to 6 geographical rotations. That is, $\mathbf{e}_6$ is calculated based on the aggregated ENS after 6 rotations, assuming that each time a different slice is affected.[12] From table VI.1, it is clear that, under the given assumptions, inequality decreases if load shedding is applied more often. Rotation between the different slices implies that those consumers who have been treated very unfairly with the first action receive a favorable treatment in the next one. However, because a large

[11]These assumptions are a simplification of the real situation in order to obtain an illustrative case study. In reality, consumers in different slices are more diversified and more subgroups can be considered in the different slices, especially in the unaffected slice 7.

[12]Power demand and load curtailment are assumed to last for a fixed time period Δt, i.e. $D_{E,i} = P_{D,i} \cdot \Delta t$ with $P_{D,i}$ the power demand of consumer i [MW] and $ENS_i = LC_i \cdot \Delta t$ with LC_i the load curtailment of consumer i [MW].

share of demand remains unaffected (slices 7), inequality is still high even after shedding each of the 6 slices once.[13]

Table VI.1: U_{ENS} after 1 up to 6 rotations of the controlled load-shedding plan proposed in Fig. VI.2.

Slice i	$\mathbf{d}$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$
1	0.04	1	0.5	0.33	0.25	0.2	0.17
2	0.04	0	0.5	0.33	0.25	0.2	0.17
3	0.04	0	0	0.33	0.25	0.2	0.17
4	0.04	0	0	0	0.25	0.2	0.17
5	0.04	0	0	0	0	0.2	0.17
6	0.04	0	0	0	0	0	0.17
7	0.76	0	0	0	0	0	0
U_{ENS}		0.96	0.92	0.88	0.84	0.80	0.76

The main concern of consumers might be their inequality compared to consumers with the same characteristics. The inequality indicator for this case is shown in table VI.2 and can be obtained by repeating the calculation in table VI.1, omitting the critical high-VOLL consumers of slice 7". The inequality has been reduced compared to the results in table VI.1, but only slightly due to the large share of unaffected consumers in slice 7' with similar characteristics as the ones in the affected slices 1 to 6.

Although it is difficult to obtain equality of reliability between consumers in the practical application of load-shedding plans, it is possible to distribute the economic consequences of the activation of load-shedding plans more equally over all consumers in the system. A practical measure might be to compensate affected consumers. If (part of) the economic burden is shared by all consumers, consequences of an interruption will be distributed more equally. However, the exact interruption cost per con-

[13] It should be noted that the effect on the inequality in the case studies can be considered as a marginal effect. Depending on the initial distribution of reliability, some decisions might make the overall distribution of reliability more equal. In a practical setting, the effect of decisions on the existing distribution of unreliability should be assessed. The initial distribution of reliability among consumers in the case studies in this paper is assumed to be equal.

Table VI.2: U_{ENS} after 1 up to 6 rotations of the controlled load-shedding plan proposed in Fig. VI.2 only considering consumers with similar characteristics.

Slice i	$\mathbf{d}$	$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	$\mathbf{e}_4$	$\mathbf{e}_5$	$\mathbf{e}_6$
1	0.047	1	0.5	0.333	0.25	0.2	0.167
2	0.047	0	0.5	0.333	0.25	0.2	0.167
3	0.047	0	0	0.333	0.25	0.2	0.167
4	0.047	0	0	0	0.25	0.2	0.167
5	0.047	0	0	0	0	0.2	0.167
6	0.047	0	0	0	0	0	0.167
7’	0.719	0	0	0	0	0	0
U_{ENS}		0.95	0.91	0.86	0.81	0.77	0.72

sumer is hard to determine, and the use of a fixed price might result in over- or undercompensation, depending on the consumer and the level of compensation.

Table VI.3 shows the impact of compensating affected consumers based on the amount of energy not supplied. The inequality indicator is calculated based on the inequality ratios defined in (6.6). The compensation per MWh equals a percentage of the weighted average VOLL of the affected consumers, ranging from no compensation up to a compensation equal to 100% of the weighted average VOLL. The economic burden of the compensation is shared between all consumers and is divided according to their demand share, for example through energy-based transmission tariffs.

In this illustrative case, the VOLL of the affected consumers is assumed to be equal. The inequality can significantly be reduced if a compensation scheme is put in place, even in the case of partial compensation. 100% compensation results in complete equality under these assumptions. However, if VOLL differs between consumers in the affected slices and interruptions are compensated at average VOLL, some consumers will be over-compensated, while others will be under-compensated, resulting in a remaining level of inequality between the consumers.

Table VI.3: Evolution of inequality as a function of the number of rotations and the relative amount of compensation.

	Rotations					
Compensation	1	2	3	4	5	6
0%	0.96	0.92	0.88	0.84	0.80	0.76
30%	0.67	0.64	0.62	0.59	0.56	0.53
50%	0.48	0.46	0.44	0.42	0.40	0.38
80%	0.19	0.18	0.18	0.17	0.16	0.15
100%	0	0	0	0	0	0

VOLL is assumed to be equal for all consumers in the affected slices.

VI.5 Case study II: Comparison of short term reliability management strategies

Due to evolutions in power systems there is a continuous search for alternative security criteria that perform better than currently used deterministic approaches. However, the currently used deterministic N-1 criterion is transparent, easy to use and has led to acceptable results in the last decades, so that power system stakeholders are not eager to change their reliability management. Evaluating performance of different reliability management approaches and criteria and comparing with the N-1 approach is crucial in order to convince power system stakeholders to move towards alternative reliability management strategies that might be more cost-effective.

Up till now, performance evaluation of reliability management was mainly based on socio-economic indicators, such as total system cost, system related indicators, such as line overloading or voltage violations, and reliability indicators, such as ENS (McCalley et al., 2004; Heylen et al., 2016b; Kirschen and Jayaweera, 2007). Next to these traditional indicators, the practical applicability of alternative reliability management strategies is also determined by their social acceptance. A fair distribution of unreliability among consumers is one of the aspects that determines the social acceptability of an alternative reliability management strategy.

VI.5.1 Data and assumptions

This case study illustrates the use of the inequality indicator U_{ENS} in a comparative study of two different short term reliability management strategies: (a) the state-of-the-art, deterministic N-1 criterion and (b) a probabilistic criterion aiming at the minimization of expected total system cost. The N-1 approach considers constant and equal VOLL for all consumers, while the probabilistic approach takes into account that VOLL differs between consumer groups and over time(Ovaere et al., 2016). VOLL data for Norway are used (EnergiNorge, 2012) and two consumer groups (residential and non-residential) are distinguished. A 5 node network, based on the Roy Billinton Reliability test system (Billinton et al., 1989), is used.[14] Detailed data about the test system and the VOLL data can be found in (Ovaere et al., 2016).

Performance evaluation of the two reliability management strategies is executed using an analytical non-sequential state enumeration technique. Two decision stages are considered in short term reliability management: day ahead operational planning and real time operation. Operational planning is simulated for both reliability management strategies under analysis. The N-1 criterion aims at securing all single branch and generator outages and the N-0 state given the forecast of net demand. All states are considered as equally probable and equally severe. The probabilistic approach on the other hand aims at minimizing the expected total system cost taking into account the most probable contingencies up to a cumulative probability of 99% and 7 possible realizations of net total demand. Operational planning is simulated for a set of time instances for which forecast values of net total demand are given.

In a second step, corrective control is simulated for a set of real-time realizations. This set is the Cartesian product of the most probable contingencies up to a cumulative probability of 99.6% and 11 possible real time realizations of net total demand derived from a normal distribution with mean equal to the forecast value of net total demand at the corresponding time instant and a standard deviation of 4%. The simulation of preventive and corrective control is executed using a DC security constrained opti-

[14]In order to serve the illustrative purpose of this case study, a 5 node test system is used. The indicator can similarly be applied to larger systems.

mal power flow (SCOPF) in which generation redispatch, branch switching, phase shifting transformer tap changing and load curtailment are considered as available actions (Van Acker and Van Hertem, 2016). The simulations are executed using a MATLAB implementation (Heylen et al., 2016b) interfacing with the DC SCOPF, which is implemented in AMPL (Fourer et al., 1987).

For simplification, the set of time instances consists of $6 \times 3 \times 4 = 72$ time instances, representing 6 periods in the year (winter, early spring, late spring, summer, early autumn and late autumn), 3 types of days (weekday, Saturday and Sunday) and 4 times during the day (night, morning, noon and evening). By weighting the outcomes of the different time instances by their probability of occurrence, we approximate the evaluation over a year (Ovaere et al., 2016).

VI.5.2 Results

TSOs and policy makers are typically interested in potential overall efficiency gains or total system cost savings when changing their reliability management. However, social acceptance is crucial in order to practically deploy alternative reliability management. Therefore, part of the selection of an alternative reliability management strategy is making a trade-off between efficiency and inequality between consumers (Ovaere et al., 2016).

Fig. VI.3 shows the Lorenz curves of inequality between consumers at different nodes ($U_{ENS,node}$) for both reliability management strategies. This figure clearly shows that inequality is higher with a probabilistic reliability criterion ($U^{prob}_{ENS,node} = 0.57$) than with a deterministic N-1 reliability criterion ($U^{N-1}_{ENS,node} = 0.11$). This is because the probabilistic approach exploits the differences in VOLL between consumer groups and over time, while the N-1 does not. As a result the probabilistic approach leads not only to a higher level of inequality of reliability, but also to lower socio-economic costs (64% lower in this case study).

Part of the efficiency gains can be used to decrease public opposition to the higher inequality of reliability. Fig. VI.4 identifies the most unfairly treated nodes by plotting the inequality ratios r_i. This figure shows that consumers from nodes 4 and 5 have a disproportionately low reliability level

Figure VI.3: Lorenz curves for inequality between nodes in terms of expected energy not served for the two reliability management strategies compared to the line of equality.

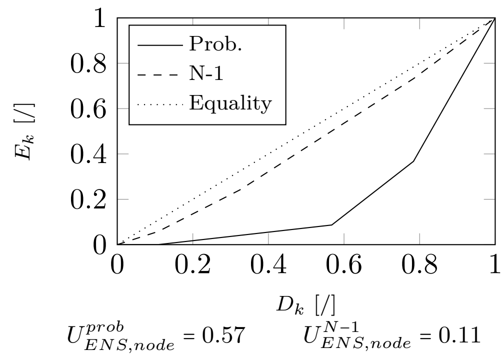
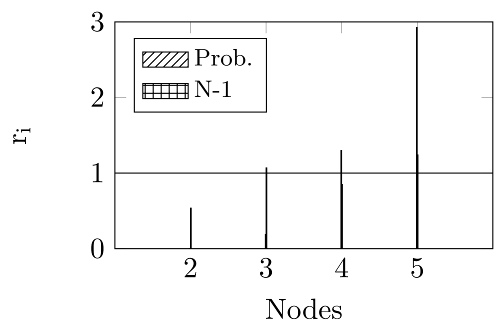


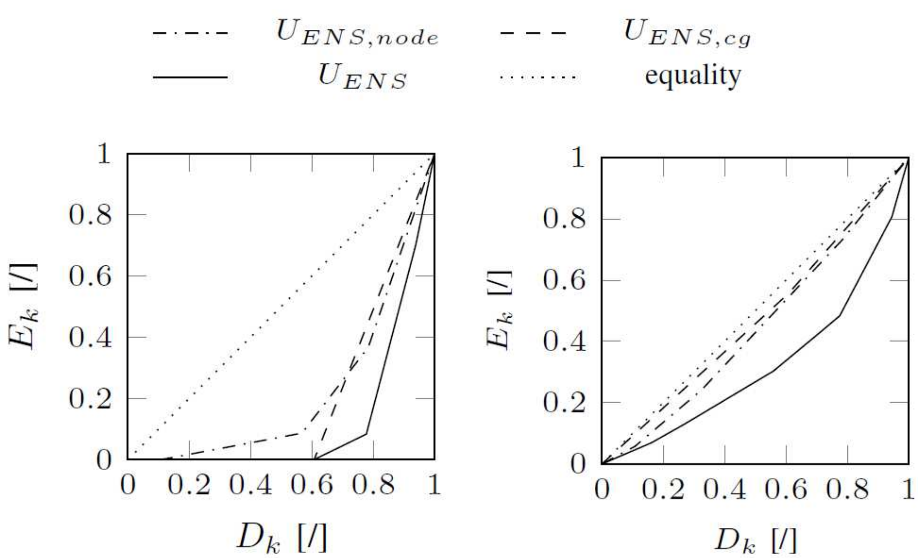
Figure VI.4: Inequality ratios per node for probabilistic reliability management and reliability management based on the N-1 criterion.



with the probabilistic approach, which means that they should be remunerated or safeguarded against other reliability-decreasing decisions.

On top of the inequality between nodes ($U_{ENS,node}$), the indicator can also be calculated for inequality between different consumer groups ($U_{ENS,cg}$) or between individual consumers (U_{ENS}). Calculating inequality between individual consumers is hard in practice, because exact energy not served and demand per consumer are not available to TSOs. They only have estimations or nodal values. However, by grouping consumers per node ($U_{ENS,node}$) or per consumer group ($U_{ENS,cg}$), the Lorenz curve is an approximation of the Lorenz curve that considers all consumers individually.[15] This is graphically illustrated in Fig. VI.5. Table VI.4 shows that this approximation of the Lorenz curve results in lower values of the inequality indicators $U_{ENS,node}$ and $U_{ENS,cg}$, quantifying the inequality between nodes and between consumer groups respectively, compared to U_{ENS}, which considers different consumer groups at different nodes. Individual inequality is always understated if aggregation is used. Nevertheless, the conclusion remains unaffected that probabilistic approaches lead to higher inequality than deterministic approaches in this case study, whatever the compared groups.

Figure VI.5: Impact of grouping consumers per node ($U_{ENS,node}$) or per consumer group ($U_{ENS,cgr}$) on the Lorenz curves



(a) Probabilistic (b) N-1

Lastly, even if data is available at the level of individual consumers, it makes sense to calculate the inequality between nodes or between consumer groups. Consumers' perception of their peers influences which groups

[15]Inequality ratios r_g per group g, i.e. per node for $U_{ENS,node}$ or per consumer group for $U_{ENS,cg}$, equal $r_g = \frac{\sum_{i \in I_g} ENS_i}{\sum_{j=1}^{n} ENS_j} \cdot \frac{\sum_{j=1}^{n} D_{E,j}}{\sum_{i \in I_g} D_{E,i}}$, with I_g the set of consumers belonging to group g.

Table VI.4: Inequality between nodes $U_{ENS,node}$, between consumer groups $U_{ENS,cg}$ and between individual consumers U_{ENS} for the two types of reliability management.

	Probabilistic	N-1
$U_{ENS,node}$	0.57	0.11
$U_{ENS,cg}$	0.61	0.05
U_{ENS}	0.75	0.35

need to be considered in the calculation of the inequality indicator. For example, Table VI.5 shows the equality indicator within groups (residential and non-residential). This table shows that for the presented case study the inequality between residential consumers does not increase much when moving from the N-1 criterion to a probabilistic criterion, while it increases more between non-residential consumers.

Table VI.5: Inequality U_{ENS} between consumers in the two considered consumer groups for the two reliability management strategies

	Consumer groups	
	Residential	Non-residential
Prob.	0.38	0.66
N-1	0.32	0.30

VI.6 Reducing inequality

If the inequality indicator shows that the distribution of unreliability among consumers is highly unequal, measures can be taken to reduce this inequality. This is possible based on the principle of transfers (Dalton, 1920). This principle states that a transfer s of the share of ENS from a consumer i to a consumer j decreases the value of the inequality indicator if $r_i > r_j$ and $\frac{e_i - s}{d_i} \geq \frac{e_j + s}{d_j}$. This requires a more detailed study to see which consumers are mostly affected in a positive and negative way. Based on this study, the TSO can decide to safeguard the most affected consumers if load curtailment is required in the future.

From a socio-economic perspective it might be better to have a certain level of inequality, e.g. in systems with remote and sparsely populated load points. In this case it is not economically viable to have the same level of redundancy for these remote load points as for a densely populated area. This decision might result in a higher share of energy not supplied in these remote load points. A cost-effective way to reduce the level of inequality in this case might be to invest in small, local generation units, possibly (partly) subsidized. Alternatively, transmission tariffs might be influenced by the local reliability level resulting in lower tariffs for consumers with a lower than average reliability level and a higher tariff for consumers with a higher than average reliability level. Other options are a market for reliability or end-consumer contracts where the price depends on the reliability level. Bi-lateral interruptible load contracts between TSOs and large industrial consumers with flexible processes are already in place nowadays, but they might be extended to include smaller consumers as well, e.g. in case of a roll-out of smart meters. These kinds of economic compensations result in a more equal distribution of the cost of unreliability, although the inequality of reliability itself is not changed. In order to obtain satisfactory results, the design of these measures should be done with care, which requires a multi-faceted analysis.

The exact determination of the interruption costs is a challenge in the compensation of affected consumers. Not only are energy not served and demand per consumer hard to obtain, also exact values of lost load per node or per consumer are rarely available in practice. However, the fourth energy package of the European Commission prescribes that all member states have to establish at least a single estimate of VOLL for their territory and can establish a VOLL per bidding zone, if they have several ones. In many other regions such obligation does not yet exist, but more and more studies are estimating VOLL with a higher level of detail, taking into account differentiation in terms of type of consumers, time and duration. An overview of these studies can be found in (Ovaere et al., 2016).

VI.7 Conclusion

Quantifying inequality of power system reliability in a single value allows stakeholders to compare the level of inequality between different entities and for different decisions, such as adequacy or security measures. An equal and fair distribution of power system unreliability is crucial to reduce public opposition and is one of the aspects to ensure social acceptance. A definition of equality based on energy not supplied is proposed. The usefulness of the indicator is illustrated in the context of load-shedding plans and in the evaluation of alternative short term reliability management strategies. In addition, measures are discussed to directly reduce inequality of reliability, such as alternating the affected customers over time, or to do so indirectly by redistributing the consequences, such as through compensation schemes, adapted transmission tariffs, markets for reliability or end-consumer contracts. Future work has to focus on the careful design of measures to reduce inequality in a cost effective way. The proposed indicator can be usefully applied in order to evaluate the impact of such measures.

Acknowledgements

We are grateful for very helpful comments from Fridrik Mar Baldursson, Guido Pepermans and Frank Verboven. Work is conducted through funding by the GARPUR project of the 7th European Community Framework Programme: http://www.garpur-project.eu/. The work of Evelyn Heylen is funded by the Research Foundation Flanders. All remaining errors are our own.

Bibliography

50Hertz Transmission GmbH, Amprion GmbH, Elia, GmbH, T. T., and TransnetBW GmbH (2014). Potential cross-border balancing cooperation between the Belgian , Dutch and German electricity Transmission System Operators. (October).

Ai, C., Martinez, S., and Sappington, D. E. M. (2003). Incentive Regulation and Customer Service Quality. *Journal of Regulatory Economics*, 26(3):263–285.

Ai, C. and Sappington, D. E. M. (2002). The Impact of State Incentive Regulation on the U . S . Telecommunications Industry. *Journal of Regulatory Economics*, 22(2):133–160.

Ajodhia, V. and Hakvoort, R. (2005). Economic regulation of quality in electricity distribution networks. *Utilities Policy*, 13:211–221.

Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, 43(6):865–880.

Armstrong, M. and Sappington, D. E. M. (2004). Toward a synthesis of models of regulatory policy design with limited information. *Journal of Regulatory Economics*, 26(1):5–21.

Arrow, K. J., Harris, T., and Marschak, J. (1951). Optimal Inventory Policy. *Econometrica*, 19(3):250–272.

Atkinson, A. (1970). On the Measurement of Income Inequality. *Journal of Economic Theory*, 2(3):244–263.

Auray, S., Mariotti, T., and Moizeau, F. (2011). Dynamic regulation of quality. *The RAND Journal of Economics*, 42(2):246–265.

Australian Energy regulator (2013). Efficiency Benefit Sharing Scheme for Electricity Network Service Providers. Technical Report November.

Averch, H. and Johnson, L. L. (1962). Behavior of the Firm Under Regulatory Constraint. *American Economic Review*, 52(5):1052–1069.

Baldursson, F. M., Lazarczyk, E., Ovaere, M., and Proost, S. (2016a). Cross-border exchange and sharing of generation reserve capacity. *CREE Working paper 14/2016*, pages 1–29.

Baldursson, F. M., Lazarczyk, E., Ovaere, M., and Proost, S. (2016b). Multi-TSO system reliability : cross-border balancing. *IEEE International Energy Conference*, pages 1–6.

Banerjee, A. (2003). Does incentive regulation 'cause' degradation of retail telephone service quality? *Information Economics and Policy*, 15(2):243–269.

Baron, D. P. and Myerson, R. B. (1982). Regulating a Monopolist with Unknown Costs. *Econometrica*, 50(4):911–930.

Bertling, L., Allan, R., and Eriksson, R. (2005). A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. *IEEE Transactions on power systems*, 20(1):75–82.

Bertsimas, D., Brown, D. B., and Caramanis, C. (2011). Theory and applications of robust optimization. *SIAM review*, 53(3):464–501.

Besanko, D., Donnenfeld, S., and White, L. J. (1987). Monopoly and Quality Distortion : Effects and Remedies. *The Quarterly Journal of Economics*, 102(4):743–768.

Billinton, R., Kumar, S., Chowdhury, N., Chu, K., Debnath, K., Goel, L., Khan, E., Kos, P., Nourbakhsh, G., and Oteng-Adjei, J. (1989). A reliability test system for educational purposes-basic data. *IEEE Trans. Power Syst.*, 4(3):1238–1244.

Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). The Nash bargaining solution in economic modelling. *The RAND Journal of Economics*, 17(2):176–188.

Blank, L., Kaserman, D. L., and Mayo, J. W. (1998). Dominant firm pricing with competitive entry and regulation: The case of intraLATA toll. *Journal of Regulatory Economics*, 14(1):35–53.

Blank, L. and Mayo, J. W. (2009). Endogenous regulatory constraints and the emergence of hybrid regulation. *Review of Industrial Organization*, 35(3):233–255.

Blumsack, S., Lave, L. B., and Ilic, M. D. (2007). A Quantitative Analysis of the Relationship Between Congestion and Reliability in Electric Power Networks. *The Energy Journal*, 28(4):73–100.

Boiteux, M. (1956). Sur la gestion des Monopoles Publics astreints a l 'equilibre budgetaire. *Econometrica*, 24(1):22–40.

Borenstein, S., Bushnell, J., and Stoft, S. (2000). The Competitive Effects of Transmission Capacity in a Deregulated Electricity Industry. *The RAND Journal of Economics*, 31(2):294.

Braess, D. (1968). Über ein Paradoxon aus der Verkehrsplanung. *Mathematical Methods of Operations Research*, 12(1):258–268.

CAISO (2016). Quantifying EIM Benefits, 2016 Q1 Report. Technical report.

Calvo, J. L., Tindemans, S. H., and Strbac, G. (2016). Incorporating failures of system protection schemes into power system operation. *Sustainable Energy, Grids and Networks*, 8:98–110.

Cambini, C., Fumagalli, E., Rondi, L., and Cambini, B. C. (2016). Incentives to quality and investment: evidence from electricity distribution in Italy. *Journal of Regulatory Economics*, 49(1):1–32.

Capitanescu, F., Fliscounakis, S., Panciatici, P., and Wehenkel, L. (2011a). Day-ahead security assessment under uncertainty relying on the combination of preventive and corrective controls to face worst-case scenarios. *PSCC proceedings*.

Capitanescu, F., Fliscounakis, S., Panciatici, P., and Wehenkel, L. (2012). Cautious operation planning under uncertainties. *IEEE Tran. Power Syst.*, 27(4):1859–1869.

Capitanescu, F., Glavic, M., Ernst, D., and Wehenkel, L. (2007). Contingency filtering techniques for preventive security-constrained optimal power flow. *IEEE Tran. Power Syst.*, 22(4):1690–1697.

Capitanescu, F., Martinez Ramos, J. L., Panciatici, P., Kirschen, D., Marano Marcolini, A., Platbrood, L., and Wehenkel, L. (2011b). State-of-the-art, challenges, and future trends in security constrained optimal power flow. *Electric Power Systems Research*, 81(8):1731–1741.

Carlsson, F. and Martinsson, P. (2008). Does it matter when a power outage occurs? - A choice experiment study on the willingness to pay to avoid power outages. *Energy Economics*, 30(3):1232–1245.

Castro, R., Faias, S., and Esteves, J. (2016). The cost of electricity interruptions in Portugal: Valuing lost load by applying the production-function approach. *Utilities Policy*, 40:48–57.

CEER (2010). Guidelines of Good Practice on Estimation of Costs due to Electricity Interruptions and Voltage Disturbances. Technical report.

CEER (2011). 5th CEER Benchmarking Report on Quality of Electricity Supply 2011. Technical report.

Cepin, M. (2011). *Assessment of power system reliability*. springer.

Chao, H.-p. (1983). Peak Load Pricing and Capacity Planning with Demand and Supply Uncertainty. *Bell Journal of Economics*, 14(1):179–190.

Chao, H.-P. and Wilson, R. (1987). Priority Service: Pricing , Investment , and Market Organization. *American Economic Review*, 77(5):899–916.

Cohen, J., Moeltner, K., Reichl, J., and Schmidthaler, M. (2016). An empirical analysis of local opposition to new transmission lines across the EU-27. *The Energy Journal*, 37(3):59–82.

Coolen, F. P., Troffaes, M. C., and Augustin, T. (2011). Imprecise probability. In *International encyclopedia of statistical science*, pages 645–648. Springer.

CRA International (2008). Assessment of the Value of Customer Reliability. *VENCorp*.

Crew, M. A. and Kleindorfer, P. R. (1976). Peak load pricing with a diverse technology. *The Bell Journal of Economics*, pages 207–231.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348–361.

de Nooij, M., Baarsma, B., Bloemhof, G., Slootweg, H., and Dijk, H. (2010). Development and application of a cost-benefit framework for energy reliability using probabilistic methods in network planning and regulation to enhance social welfare: The N-1 rule. *Energy Economics*, 32(6):1277–1282.

de Nooij, M., Koopmans, C., and Bijvoet, C. (2007). The value of supply security. The costs of power interruptions: Economic input for damage reduction and investment in networks. *Energy Economics*, 29(2):277–295.

de Nooij, M., Lieshout, R., and Koopmans, C. (2009). Optimal blackouts: Empirical results on reducing the social cost of electricity outages through efficient regional rationing. *Energy Economics*, 31(3):342–347.

Decker, C. (2014). *Modern Economic Regulation: : An introduction to theory and practice*. Cambridge University Press.

Dijk, J. and Willems, B. (2011). The effect of counter-trading on competition in electricity markets. *Energy Policy*, 39(3):1764–1773.

Dogan, G., Labeau, P.-E., Maun, J.-C., Sprooten, J., Galvez, M., and Sleurs, K. (2016). The discrete forecast error scenarios method for grid reliability assessment in short-term planning under renewable generation and cross-border flows. In Walls, L., Revie, M., and Bedford, T., editors, *Risk, Reliability and Safety: Innovating Theory and Practice*, Glasgow.

Doorman, G. L. and Van der Veen, R. (2013). An analysis of design options for markets for cross-border balancing of electricity. *Utilities Policy*, 27(January 2011):39–48.

Doucet, J., Kleit, A., and Fikirdanis, S. (2013). Valuing electricity transmission: The case of Alberta. *Energy Economics*, 36:396–404.

Ela, E., Milligan, M., and Kirby, B. (2011). Operating Reserves and Variable Generation. Technical Report August.

Electricity Authority, N. Z. (2013). Investigation into the value of lost load in New Zealand - Report on methodology and key findings. (July):1–87.

Elia (2016). Bevoorradingszekerheidstudie voor België. Technical report.

EnergiNorge (2012). *Samfunnsøkonomiske Kostnader Ved Avbrudd Og Spenningsforstyrrelser*. Number 349.

ENTSO-E (2013). Network Code on Load-Frequency Control and Reserves. Technical report.

ENTSO-E (2014a). 10-Year Network Development Plan 2014. Technical report.

ENTSO-E (2014b). Network Code on Electricity Balancing. Technical report.

ENTSO-E (2016a). Commission regulation establishing a network code on electricity emergency and restoration (pending). Technical report.

ENTSO-E (Accessed: 1st of August 2016b). Public consultation for the design of the terre. *https://consultations.entsoe.eu/markets/terre*.

Estache, A., Rossi, M. A., and Ruzzier, C. A. (2004). The case for international coordination of electricity regulation: Evidence from the measurement of efficiency in South America. *Journal of Regulatory Economics*, 25(3):271–295.

European Commission (2015). Achieving the 10% electricity interconnection target.

European Commission (2016). Commission Regulation (EU) No .../2016 of XXX establishing a guideline on electricity transmission system operation.

Farahmand, H., Aigner, T., Doorman, G. L., Korpas, M., and Huertas-Hernando, D. (2012). Balancing market integration in the northern european continent: A 2030 case study. *IEEE Transactions on Sustainable Energy*, 3(4):918–930.

FERC (2006). Order No. 679. Technical report.

Fourer, R., Gay, D. M., and Kernighan, B. W. (1987). *AMPL: A mathematical programming language*. AT&T Bell Laboratories Murray Hill, NJ 07974.

Fu, W. and McCalley, J. D. (2001). Risk based optimal power flow. In *Power Tech Proceedings, 2001*. IEEE.

GARPUR (2014). D1.2 Current practices, drivers and barriers for new reliability standards. Technical report.

GARPUR (2015). D6.1 Functional analysis of System Operation processes. Technical report.

GARPUR (2016). Quantification method in the absence of market response and with market response taken into account. Technical report.

Gasmi, F., Ivaldi, M., and Laffont, J.-J. (1994). Rent extraction and incentives for efficiency in recent regulatory proposals. *Journal of Regulatory Economics*, 6:151–176.

Gebrekiros, Y. T., Doorman, G. L., Farahmand, H., and Jaehnert, S. (2013). Benefits of cross-border reserve procurement based on pre-allocation of transmission capacity. In *PowerTech (POWERTECH), 2013 IEEE Grenoble*, pages 1–6. IEEE.

Giannakis, D., Jamasb, T., and Pollitt, M. G. (2005). Benchmarking and incentive regulation of quality of service: An application to the UK electricity distribution networks. *Energy Policy*, 33:2256–2271.

Glachant, J. M. and Pignon, V. (2005). Nordic congestion's arrangement as a model for Europe? Physical constraints vs. economic incentives. *Utilities Policy*, 13(2 SPEC. ISS.):153–162.

Grigg, C., Wong, P., Albrecht, P., Allan, R., Bhavaraju, M., Billinton, R., Chen, Q., Fong, C., Haddad, S., and Kuruganty, S. (1999). The IEEE reliability test system-1996. a report prepared by the reliability test system task force of the application of probability methods subcommittee. *IEEE Transactions on Power Systems*, 14(3):1010–1020.

Growitsch, C., Malischek, R., Sebastian, N., and Wetzel, H. (2013). The Costs of Power Interruptions in Germany - an Assessment in the Light of the Energiewende.

Hasenheit, M. (2014). Inequality: Where palma is better than gini. [Online] `http://netgreen-project.eu/blog/2014/06/18/inequality-where-palma-better-gini`.

Hauge, J. and Sappington, D. (2010). *Pricing in Network Industries*.

He, J., Cheng, L., Kirschen, D. S., and Sun, Y. (2010). Optimising the balance between security and economy on a probabilistic basis. *IET generation, transmission & distribution*, 4(12):1275–1287.

Hesseling, D. and Sari, M. (2006). The introduction of quality regulation of electricity distribution in The Netherlands. *European Energy Law Report III*, pages 127–145.

Heylen, E., Deconinck, G., and Van Hertem, D. (2016a). Analysis framework for performance evaluation of reliability management in power systems with increased uncertainty. In Walls, L., Revie, M., and Bedford, T., editors, *Risk, Reliability and Safety: Innovating Theory and Practice*, pages 2322 – 2329. Taylor & Francis Group.

Heylen, E., Labeeuw, W., Deconinck, G., and Hertem, D. V. H. (2016b). Framework for evaluating and comparing performance of power system reliability criteria. *IEEE Transactions on Power Systems*, 31(6):5153–5162.

Heylen, E., Ovaere, M., Deconinck, G., and Van Hertem, D. (2017a). An inequality indicator of power system reliability. *IEEE Tran. Power Syst. (Submitted)*.

Heylen, E., Ovaere, M., Proost, S., Deconinck, G., and Hertem, D. V. (2017b). A Multi-Dimensional Analysis of Reliability Criteria: From Deterministic N-1 to a Probabilistic Approach.

Hobbs, B. F., Rijkers, F. A., and Boots, M. G. (2005). The More Cooperation, the More Competition? A Cournot Analysis of the Benefits of Electric Market Coupling. *The Energy Journal*, 26(4):69–97.

Hofmann, M., Slejeseth, H., Volden, G. H., and Kjolle, G. (2010). Study on Estimation of Costs due to Electricity Interrumptions and Voltage Disturbances. *SINTEF Energy Research*, (December):1–72.

Hogan, W. W., Rosellón, J., and Vogelsang, I. (2010). Toward a combined merchant-regulatory mechanism for electricity transmission expansion. *Journal of Regulatory Economics*, 38(2):113–143.

Holmberg, P. and Lazarczyk, E. (2015). Comparison of congestion management techniques: Nodal, zonal and discriminatory pricing. *The Energy Journal*, 36(2):1–34.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 55(2):303–328.

Joskow, P. L. (2005). Transmission policy in the United States. *Utilities Policy*, 13(2):95–115.

Joskow, P. L. (2006). Patterns of Transmission Investment. In *Competitive Electricity Markets and Sustainability*, pages 169–235.

Joskow, P. L. (2008a). Incentive Regulation and Its Application to Electricity Networks. *Review of Network Economics*, 7(4):547–560.

Joskow, P. L. (2008b). Lessons Learned from Electricity Market Liberalization.

Joskow, P. L. (2014). Incentive Regulation in Theory and Practice: Electricity Distribution and Transmission Networks. In Nancy L. Rose, editor, *Economic Regulation and Its Reform: What Have We Learned?*, pages 291–344. University of Chicago Press Volume.

Joskow, P. L. and Tirole, J. (2007). Reliability and competitive electricity markets. *The RAND Journal of Economics*, 38(1):60–84.

Just, S. and Weber, C. (2015). Strategic behavior in the German balancing energy mechanism: incentives, evidence, costs and solutions. *Journal of Regulatory Economics*, pages 218–243.

Kaestner, R. and Kahn, B. (1990). The effects of regulation and competition on the price of AT&T intrastate telephone service. *Journal of Regulatory Economics*, 2(4):363–377.

Karangelos, E. and Wehenkel, L. (2016a). Probabilistic Reliability Management Approach and Criteria for Power System Real-time Operation. *Power Systems Computation Conference (PSCC) 2016*, (608540).

Karangelos, E. and Wehenkel, L. (2016b). Probabilistic reliability management approach and criteria for power system real-time operation. In *Power Systems Computation Conference*, pages 1–9.

Kidokoro, Y. (2002). The effects of regulatory reform on quality. *Journal of the Japanese and International Economies*, 16:135–146.

Kirschen, D. (2002). Power system security. *Power Engineering Journal*, 16(5):241–248.

Kirschen, D. and Jayaweera, D. (2007). Comparison of risk-based and deterministic security assessments. *IET Generation, Transmission & Distribution*, 1(4):527.

Kirschen, D. and Strbac, G. (2004a). *Fundamentals of Power System Economics.* John Wiley & Sons.

Kirschen, D. and Strbac, G. (2004b). Why investments do not prevent blackouts. *The Electricity Journal*, 17(2):29–36.

Kjolle, G., Samdal, K., Singh, B., and Kvitastein, O. a. (2008). Customer costs related to interruptions and voltage problems: Methodology and results. *IEEE Transactions on Power Systems*, 23(3):1030–1038.

Kleindorfer, P. R. and Fernando, C. S. (1993). Peak-Load Pricing and Reliability Under Uncertainty. *Journal of Regulatory Economics*, 5:5–23.

Kolstad, C. D. (2005). Piercing the veil of uncertainty in transboundary pollution agreements. *Environmental and Resource Economics*, 31(1):21–34.

Konow, J. (2003). Which is the fairest one of all? A positive analysis of justice theories. *Journal of Economic Literature*, 41(4):1188–1239.

Laffont, J.-J. (1994). The new economics of regulation ten years after. *Econometrica*, 62(3):507–537.

Laffont, J.-J. and Tirole, J. (1986). Using Cost Observation to Regulate Firms. *Journal of Political Economy*, 94(3):614–641.

Laffont, J.-J. and Tirole, J. (1991). Provision of quality and power of incentive schemes in regulated industries.

Laffont, J.-J. and Tirole, J. (1993). *A Theory of Incentives in Procurement and Regulation.*

Leahy, E. and Tol, R. S. J. (2011). An estimate of the value of lost load for Ireland. *Energy Policy*, 39(3):1514–1520.

Léautier, T.-O. (2000). Regulation of an Electric Power Transmission Company. *The Energy Journal*, 21(4):61–92.

Lewis, T. R. and Sappington, D. E. M. (1991). Incentives for monitoring quality. *The RAND Journal of Economics*, 22(3):370–384.

Lewis, T. R. and Sappington, D. E. M. (1992). Incentives for Conservation and Quality-Improvement by Public Utilities. *American Economic Review*, 82(5):1321–1340.

Linares, P. and Rey, L. (2013). The costs of electricity interruptions in Spain: Are we sending the right signals? *Energy Policy*, 61:751–760.

Liston, C. (1993). Price-cap versus Rate-of-Return Regulation. *Journal of Regulatory Economics*, 5(1):25–48.

Littlechild, S. (1983). Regulation of British Telecommunications' profitability: report to the Secretary of State. Technical report, Department of Industry.

London Economics (2013). The Value of Lost Load (VoLL) for Electricity in Great Britain: Final report for OFGEM and DECC. (July):1–225.

Lyon, T. P. (1996). A model of sliding-scale regulation. *Journal of Regulatory Economics*, 9(3):227–247.

Mathios, A. D. and Rogers, R. P. (1989). The impact of alternative forms of state regulation of AT & T on direct-dial, long-distance telephone rates. *The RAND Journal of Economics*, 20(3):437–453.

McCalley, J., Asgarpoor, S., Bertling, L., Billinton, R., Chao, H., Chen, J., Endrenyi, J., Fletcher, R., Ford, A., Grigg, C., Hamoud, G., Logan, D., Meliopoulous, A., Ni, M., Rau, N., Salvaderi, L., Schilling, M., Schlumberger, Y., Schneider, A., and Singh, C. (2004). Probabilistic security assessment for power system operations. *IEEE Power Engineering Society General Meeting*.

Mitchell, R. C. and Carson, R. T. (1989). *Using surveys to value public goods: the contingent valuation method*. Resources for the Future, Washington DC.

Mott MacDonald (2013). Impact Assessment on European Electricity Balancing Market. (March):1–135.

Munasinghe, M. (1981). Optimal electricity supply. *Energy Economics*, 3(3):140–152.

Munasinghe, M. and Gellerson, M. (1979). Economic criteria for optimizing power system reliability levels. *The Bell Journal of Economics*, pages 353–365.

Nash, J. F. (1953). Two-Person Cooperative Games. *Econometrica*, 21(1):128–140.

NERC (2007). Definition of ”adequate level of reliability”. Available [Online]: `http://www.nerc.com/docs/pc/Definition-of-ALR-approved-at-Dec-07-OC-PC-mtgs.pdf`.

NERC (2011a). Automatic underfrequency load shedding. Standard prc-006-1.

NERC (2011b). Load shedding plans. Standard eop-003-2.

NERC (2015). State of Reliability 2015. Technical report.

Neuhoff, K., Barquin, J., Bialek, J. W., Boyd, R., Dent, C. J., Echavarren, F., Grau, T., von Hirschhausen, C., Hobbs, B. F., Kunz, F., Nabe, C., Papaefthymiou, G., Weber, C., and Weigt, H. (2013). Renewable electric energy integration: Quantifying the value of design of markets for international transmission capacity. *Energy Economics*, 40:760–772.

Neuhoff, K. and Richstein, J. (2016). Coordinated Balancing of the European Power System.

Newbery, D. and Grubb, M. (2014). The Final Hurdle?: Security of supply, the Capacity Mechanism and the role of interconnectors.

NordREG (2012). Economic regulation of TSOs in the Nordic countries. Technical report.

NREL (2011). Flexibility Reserve Reductions from an Energy Imbalance Market with High Levels of Wind Energy in the Western Interconnection. Technical Report October.

OECD (2016). Purchasing power parities (ppp) statistics. [Online]: `http://stats.oecd.org/Index.aspx?DataSetCode=PPPGDP` Accessed on 23 August 2016.

OFGEM (2016). RIIO-ET1 Annual Report 2015-2016. Technical report.

OFGEM (2017a). Energy spend as a percentage of total household expenditure (UK).

OFGEM (2017b). RIIO-GD1 Annual Report 2015-2016. Technical report.

OFWAT (2013). Appendix 1 : Integrating the calibration of outcome delivery and cost performance incentives. Technical report.

OFWAT (2014). Setting price controls for 2015-20: Final price control determination notice : policy chapter A3 â€“ wholesale water and wastewater. Technical report.

Ovaere, M., Heylen, E., Proost, S., Deconinck, G., and Van Hertem, D. (2016). How detailed value of lost load data impact power system reliability decisions: a trade-off between efficiency and equality. *KU Leuven Department of Economics Discussion Paper Series*, 16.26.

Ovaere, M. and Proost, S. (2016). Electricity transmission reliability: The impact of reliability criteria. *KU Leuven Department of Economics Discussion Paper Series*, 16.21.

Pepermans, G. (2011). The value of continuous power supply for Flemish households. *Energy Policy*, 39(12):7853–7864.

Pérez-Arriaga, I. J., Rubio, F., Puerta, J., Arceluz, J., and Marin, J. (1995). Marginal pricing of transmission services: an analysis of cost recovery. *IEEE Transactions on Power Systems*, 10(1):546–553.

Platbrood, L., Crisciu, H., Capitanescu, F., and Wehenkel, L. (2011). Solving very large-scale security-constrained optimal power flow problems by combining iterative contingency selection and network compression. In *PSCC conference*.

Pozo, D., Contreras, J., and Sauma, E. E. (2013). If you build it, he will come: Anticipative power transmission planning. *Energy Economics*, 36:135–146.

Promislow, S. D. (1987). Measurement of equity. *Transactions of the Society of Actuaries*, 39:215–256.

Reichl, J., Kollmann, A., Tichler, R., and Schneider, F. (2008). The importance of incorporating reliability of supply criteria in a regulatory system of electricity distribution: An empirical analysis for Austria. *Energy Policy*, 36(10):3862–3871.

Reichl, J., Schmidthaler, M., and Schneider, F. (2013). The value of supply security: The costs of power outages to Austrian households, firms and the public sector. *Energy Economics*, 36:256–261.

Rosellón, J. (2007). An incentive mechanism for electricity transmission expansion in Mexico. *Energy Policy*, 35(5):3003–3014.

Rosellón, J. and Weigt, H. (2011). A Dynamic Incentive Mechanism for Transmission Expansion in Eletricity Networks: Theory, Modeling, and Application. *The Energy Journal*, Vol. 32(June).

Sappington, D. E. M. (2002). *Price Regulation and Incentives.*

Sappington, D. E. M. (2003). The Effects of Incentive Regulation on Retail Telephone Service Quality in the United States. *Review of Network Economics*, 2(4):355–375.

Sappington, D. E. M. (2005). Regulating Service Quality : A Survey. *Journal of Regulatory Economics*, 27(2):123–154.

Sappington, D. E. M. and Weisman, D. L. (2010). *Price cap regulation: What have we learned from 25 years of experience in the telecommunications industry?*, volume 38.

Schleifer, A. (1985). A Theory of Yardstick Competition. *The RAND Journal of Economics*, 16(3):385–395.

Schmalensee, R. (1989). Good Regulatory Regimes. *The RAND Journal of Economics*, 20(3):417–436.

Schmidthaler, M., Cohen, J., Reichl, J., and Schmidinger, S. (2015). The effects of network regulation on electricity supply security: a European analysis. *Journal of Regulatory Economics*, (0401).

Schutz, R. (1951). On measurement of income inequality. *American Economic Review*, 41:107–122.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Journal of the Econometric Society*, pages 219–231.

Sheshinski, E. (1976). Price, Quality and Quantity Regulation in Monopoly Situations. *Economica*, 43(170):127–137.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. *Econometrica*, 48(3):613–625.

Spence, M. (1975). Monopoly, Quality, and Regulation. *Bell Journal of Economics*, 6(2):417–429.

Steiner, P. (1957). Peak Loads and Efficient Pricing. *The Quarterly Journal of Economics*, 71(4):585–610.

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell Journal of Economics*, 2(1):3–21.

Strbac, G., Ahmed, S., Kirschen, D., and Allan, R. (1998). A method for computing the value of corrective security. *IEEE Transactions on Power Systems*, 13(3):1096–1102.

Sullivan, M. J. and Keane, M. (1995). Outage cost estimation guidebook. *Electric Power Research Institute.*

Sullivan, M. J., Mercurio, M., and Schellenberg, J. (2009). Estimated Value of Service Reliability for Electric Utility Customers in the United States. *Lawrence Berkeley National Laboratory.*

Tangerås, T. P. (2012). Optimal transmission regulation of an integrated energy market. *Energy Economics*, 34(5):1644–1655.

Ter-Martirosyan, A. and Kwoka, J. (2010). Incentive regulation, service quality, and standards in U.S. electricity distribution. *Journal of Regulatory Economics*, 38(3):258–273.

Theil, H. (1967). *Economics and Information Theory.* North Holland.

Transpower (2013). Ancillary services procurement plan. Technical Report December.

Turvey, R. (2006). Interconnector economics. *Energy Policy*, 34(13):1457–1472.

Vadlamudi, V. V., Hamon, C., Kjølle, G., and Perkin, S. (2016). On improving data and models on corrective control failures for use in probabilistic reliability management. In *Probabilistic Methods Applied to Power Systems (PMAPS), 2016 International Conference on*, pages 1–6. IEEE.

Van Acker, T. and Van Hertem, D. (2016). Linear representation of preventive and corrective actions in opf models. In *Young researchers symposium*. IEEE IAS/PELS/PES Benelux Chapter.

Van den Bergh, K., Delarue, E., and D'haeseleer, W. (2014). DC power flow in unit commitment models.

Van den Bergh, K., Hytowitz, R. B., Delarue, E., William, D., and Hobbs, B. (2017). Benefits of coordinating sizing, allocation and activation of reserves among market zones. *Electric Power Systems Research*, 143:140–148.

van der Weijde, A. H. and Hobbs, B. F. (2011). Locational-based coupling of electricity markets: benefits from coordinating unit commitment and balancing markets. *Journal of Regulatory Economics*, 39(3):223–251.

van der Weijde, A. H. and Hobbs, B. F. (2012). The economics of planning electricity transmission to accommodate renewables: Using two-stage optimisation to evaluate flexibility and the cost of disregarding uncertainty. *Energy Economics*, 34(6):2089–2101.

Vandezande, L., Saguan, M., Meeus, L., Glachant, J. M., and Belmans, R. (2009). Assessment of the implementation of cross-border balancing trade between Belgium and the Netherlands. *2009 6th International Conference on the European Energy Market, EEM 2009*, pages 1–6.

Vinod, T., Yan, W., and Xibo, F. (2001). Measuring education inequality: Gini coefficients of education. *World Bank Policy Research Working Paper 2525*.

Wang, Q. and McCalley, J. D. (2013). Risk and n-1 criteria coordination for real-time operations. *IEEE Transactions on Power Systems*, 28(3):3505–3506.

Weisman, D. L. (2005). Price regulation and quality. *Information Economics and Policy*, 17:165–174.

Williamson, O. E. (1966). Peak-Load Pricing and Optimal Capacity under Indivisibility Constraints. *American Economic Review*, 56(4):810–827.

Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., De Vos, P., Verstraete, W., and Boon, N. (2009). Initial community evenness favours functionality under selective stress. *Nature*, 458:623–626.

Zachariadis, T. and Poullikkas, A. (2012). The costs of power outages: A case study from Cyprus. *Energy Policy*, 51:630–641.

Zhou, Z., Levin, T., and Conzelmann, G. (2016). Survey of U.S. Ancillary Services Markets. Technical report, Center for Energy, Environmental, and Economic Systems Analysis, Energy Systems Division, Argonne National Laboratory.

Zimmerman, R. D., Murillo-Sanchez, C. E., and Thomas, R. J. (2011). Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on Power Systems*, 1:12–19.

List of Figures

List of Tables

A list of doctoral dissertations from the Faculty of Business and Economics can be found at the following website:

http://www.kuleuven.be/doctoraatsverdediging/archief.htm.